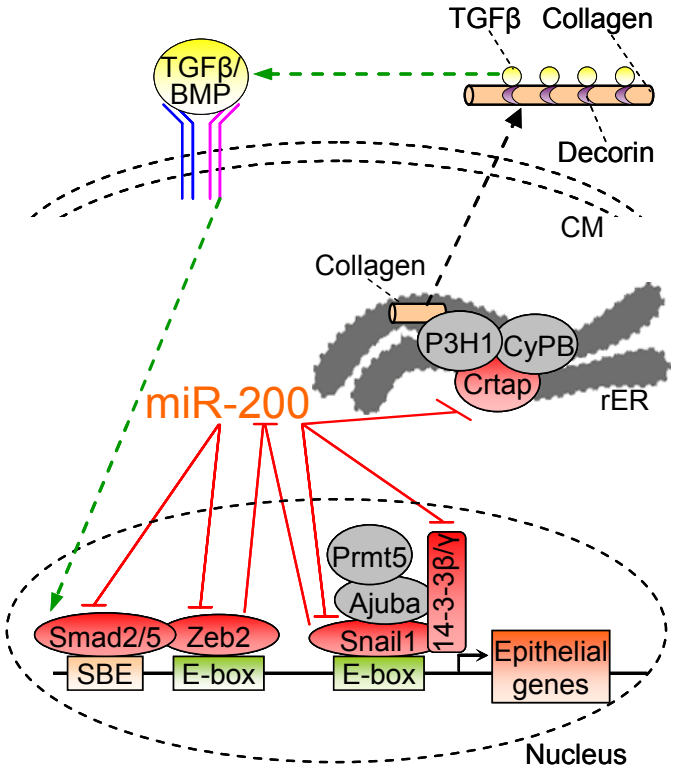
The role of the miR-200 family in the epithelial-to-mesenchymal transition (EMT)



Ricardo Perdigão Henriques

Dissertation presented to obtain a Ph.D. degree in Biochemistry, Molecular Biology at the Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa

Oeiras, August 2015

The role of the miR-200 family in the epithelial-to-mesenchymal transition (EMT)

Ricardo Perdigão Henriques

Dissertation presented to obtain a Ph.D. degree in Biochemistry, Molecular Biology at the Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa

Supervisors:

Prof. Judy Lieberman and Prof. Manuel Carrondo

Instituto de Tecnologia Química e Biológica

Universidade Nova de Lisboa

Oeiras, August 2015

The role of the miR-200 family in the epithelial-to-mesenchymal transition (EMT)

by Ricardo Perdigão Henriques

First edition: August 2015

Copyright number:

Prof. Judy Lieberman's Laboratory
Program in Cellular and Molecular Medicine
Boston Children's Hospital
Harvard Medical School
200 Longwood Ave, WAB 255
Boston MA 02115
http://labs.idi.harvard.edu/lieberman/

ITQB-UNL/IBET Animal Cell Technology Unit
Instituto de Tecnologia Química e Biológica-Universidade Nova de Lisboa/
Instituto de Biologia Experimental e Tecnológica
Av. da República EAN, 2780-157 Oeiras, Portugal
Fax: +351 21 442 11 61; Phone: +351 21 446 91 00
http://www.itqb.unl.pt
http://www.ibet.pt

Supervisors

Prof. Dr. Judy Lieberman, Chair in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, USA and Professor of Pediatrics, Harvard Medical School, Boston, MA, USA.

Prof. Dr. Manuel Carrondo, Cathedratic Professor of Chemical and Biochemical Engineering at Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal and IBET Executive Director, Oeiras, Portugal.

Jury

Dr. Luís Moita, Principal Investigator and Head of the Innate Immunity and Inflammation laboratory at Instituto Gulbenkian de Ciência, Oeiras, Portugal.

Prof. Dr. Paula Alves, Principal Investigator and Head of the Animal Cell Technology laboratory at ITQB, Oeiras, Portugal and IBET CEO, Oeiras, Portugal.

Dr. Philippe Couttet, Head of the Pathway Integration laboratory at Novartis Pharma AG, Basel, Switzerland.

Dr. Sónia Melo, Principal Investigator at IPATIMUP, Porto, Portugal.

Foreword

The present thesis dissertation is the result of six years of research at Prof. Judy Lieberman's laboratory at the Harvard Medical School.

This thesis work intends to provide the first high-confidence genome-wide map of direct miR-200c target genes and to study the biological role of these novel target genes.

Acknowledgements

I would like to express my deepest gratitude to everyone that supported me during my doctoral studies. Innumerous people contributed directly or indirectly to the work within this dissertation.

I would like to thank my advisor Prof. Manuel Carrondo for accepting me as his doctoral student and for all his guidance and moral support during all these years. He was critical to my move to the USA, as he offered me the opportunity to work at MIT in a collaboration with Prof. Daniel Wang. This was an enormous step in my career that subsequently opened many opportunities. For this I will always be grateful and will never forget his support.

I would also like to thank my advisor Prof. Judy Lieberman for accepting me as her student in her world-renowned research group, for her mentorship and financial support over many years. Her support and guidance allowed me to pursue interesting research questions with great scientific rigor. I thank her for always being willing to take time out of her busy schedule to meet with me and offer her scientific insights.

I also want to thank Prof. Paula Alves for always supporting my career and offering me her guidance from the time I was an intern student at her laboratory. She has been a great mentor and friend.

I am also grateful to all Lieberman lab members for years of advice, support and friendship. Specially, I would like to thank Dr. Fabio Petrocca, who taught me all basic molecular biology techniques and nurtured my scientific interests. His patience and knowledge allowed me to rapidly transition from the bioengineering field to the molecular oncology field. I will always remember his support during difficult times, and it was an honor to work with him in this project. Thank you for being a remarkable scientist and person!

I also thank Prof. Guilherme Ferreira (UALG) and Prof. João Crespo (UNL) for accepting to be part of my thesis committee.

A special thanks to my thesis defense jury members, Dr. Luís Moita, Prof. Paula Alves, Dr. Philippe Couttet and Dr. Sónia Melo for taking the time to evaluate my work.

I would also like to give a special thanks to Gabriel Altschuler, Marshall Thomas, Minh Le, Shen Mynn Tan, Linfeng Huang and Francisco Navarro for always offering me their help which was critical to pursue and publish my research.

I would also like to thank current and previous members of the Lieberman laboratory Radiana Trifonova, Zhan Xu, Michael Walch, Ashish Lal, Nan Yan, Anders Wittrup, Adi Gilboa-Geoffen, Elizabeth O'Day, Gregory Idos, Emre Basar, Jerome Thiery, Danielle Rajani, Dennis Martinvalet, Laura Maliszewski, Dipanjan Chowdhury, Farokh Dotiwala, Changying Guo, Sachin Mulkin, Brooke Bollman and Xing Liu for all their support and friendship during my six year stay in the lab.

I also thank my friends at MIT Raga Markely, David McClain, Sohan Patel and Andrés Abin for their support during my work there.

I also express my deepest gratitude to my friends in Boston that helped me survive the cold and long New England winters, Pedro Caetano, Jonathan Soto, Elvira Rocha, Francisco Queiró, Jorge Fradinho, Pedro Pinto, Marcus Dahlem, Priyadarshi Panda, Sigurd Østrem and Luca Pinto.

I would also like to thank current and previous members of the Animal Cell Technology laboratory that have supported me since I met them in 2006: António Roldão, Marcos Sousa, Tiago Ferreira, Margarida Serra, Nuno Carinhas, Ana Teixeira, Ana Coroadinha, Catarina Brito, Cristina Peixoto, Carina Silva and Pedro Cruz.

I also thank my longtime friends in Portugal, that even at a great distance always gave me their support and never forgot me: Bruno Vilares, Dr. Adelino Cunha, João Carreiro, Hugo Botelho and Rita Vaz.

I also thank the Portuguese Ministry of Science and Technology (FCT) for my Ph.D. fellowship SFRH/BD/37188/2007.

Um sentido agradecimento ao meu grande amigo Bruno Vilares, foi crítico o teu apoio durantes todos estes anos de intensa amizade!

Um especial agradecimento para ti Adiari, companheira de vida e de luta há 8 anos! O teu Amor fez-me Sonhar mais Alto e alcançar muito Mais... Obrigado por tudo!

Agradeço também ao Quim Perdigão, Tina, Juka, Teresa Sampaio e Família Vázquez Rodríguez todo o apoio que não esquecerei!

Queria terminar com um importante agradecimento a toda a minha Família, especialmente à minha mãe Luísa e avós Jacinta e Manuel pela crença que sempre tiveram nas minhas capacidades e Amor desde sempre!

Ph.D. Publications

Perdigão-Henriques R, Petrocca F, Altschuler G, Thomas MP, Le MT, Tan SM, Hide W, Lieberman J. miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes. *Oncogene*. Epub Mar 23, 2015.

Le MT, Hamar P, Guo C, Basar E, **Perdigão-Henriques R**, Balaj L, Lieberman J. miR-200-containing extracellular vesicles promote breast cancer cell metastasis. *J Clin Invest*. 2014; 124(12):5109-28.

Publications in conference proceedings

Perdigão-Henriques R, *et al.* (2009, 2010, 2011, 2012 and 2013). The role of miR-200 in breast cancer metastasis, Annual Harvard Medical School-Immune Disease Institute Retreat, Boston, USA.

Perdigão-Henriques R, *et al.* (2011). The role of miR-200 in breast cancer metastasis, 6th Microsymposium on small RNAs, Vienna, Austria.

Perdigão-Henriques R, *et al.* (2013). The role of miR-200 in breast cancer metastasis, Keystone Symposium on Noncoding RNAs in Development and Cancer, Vancouver, Canada.

Dissertation Abstract

Over the last decade, microRNAs (miRNAs) have earned a lot of attention due to their critical roles in several biological processes and human diseases. However, progress in this field has been limited by the difficulty of discovering miRNA target genes, as each miRNA can potentially bind to hundreds of different mRNAs. Therefore, the main focus of this thesis was to develop an integrated framework that could be used to discover novel genes and pathways regulated by a given miRNA. We applied this framework to study the miR-200 family, which was previously shown to be involved in embryonic development and tumor metastasis but only a few target genes were known.

In **Chapter 1**, miRNA biogenesis and action mechanism are described. miRNA target prediction tools, both computational and experimental, are also reviewed. Also, a basic overview of the epithelial to mesenchymal transition (EMT) and TGF-β signaling is given as both are regulated by miR-200. We also provide an overview of the miR-200 family and a detailed list of published miR-200 target genes.

In **Chapter 2**, we describe the framework we developed to discover novel miR-200c-regulated genes and cellular pathways. We found 520 putative miR-200c-regulated mRNAs in mouse, most of which were novel targets not described before, as in the beginning of this project only a dozen of miR-200 targets were known. Accordingly, these putative target mRNAs were highly enriched for miR-200c seed pairing and most were down-regulated by miR-200c over-expression. We also show that this framework has a low false-positive discovery rate. Additionally, results show that miR-200c mainly regulates 13 cellular pathways. The pathways most significantly regulated are the epidermal growth factor receptor (EGFR) and the mitogen-activated protein kinase (MAPK) signaling pathways.

Moreover, we show that miR-200c regulates the transforming growth factor-beta (TGF-β) signaling pathway, among other pathways. Accordingly, we experimentally showed that miR-200c counteracts the suppressive effects of TGF-β and bone morphogenetic protein-2 (BMP-2) on epithelial gene expression. Additionally, we showed that miR-200c regulates the 3'-untranslated regions (3'UTRs) of 12 (*Crtap, Fhod1, Smad2, Map3k1, Tob1, Ywhag/14-3-3γ, Ywhab/14-3-3β, Smad5, Zfp36, Xbp1, Mapk12*, *Snail1*) of 14 putative miR-200c-binding mRNAs tested. These 12 genes were experimentally validated as miR-200c targets by identifying their 3'UTR miR-200c recognition elements. We also experimentally showed that for these 12 genes the extent of mRNA binding to miR-200c strongly correlated with gene suppression.

Smad2 and Smad5 form a complex with Zeb2 and Ywhab/14-3-3β and Ywhag/14-3-3γ form a complex with Snail1. These complexes that repress transcription assemble on epithelial gene promoters. We showed that miR-200c over-expression increases RNA polymerase II binding to epithelial gene promoters, while reducing binding of Zeb2 and Snail1 complexes to these promoters. Expression of miR-200c-resistant *Smad5* modestly, but significantly, reduced epithelial gene induction by miR-200c.

miR-200 expression or *Zeb2* knockdown are known to inhibit cell invasion in *in vitro* assays. Knockdown of each of 3 novel miR-200 target genes identified here, *Smad5*, *Ywhag* and *Crtap*, also profoundly suppressed cell invasion. Thus miR-200 suppresses TGF-β/BMP signaling, promotes epithelial gene expression and suppresses cell invasion by regulating a network of genes.

In **Chapter 3**, main conclusions and inherent limitations are discussed and future work is outlined.

In summary, this thesis provides a framework that can be used to identify novel genome-wide targets and miRNA-regulated pathways of other miRNAs to better understand their biological role. Importantly, this thesis substantially contributes to both the miR-200 and EMT fields by providing a global list of novel miR-200c-regulated targets and cellular pathways and revealing how miR-200 orchestrates EMT.

Resumo da Dissertação

Durante a última década os microRNAs (miRNAs) têm recebido muita atenção devido ao seu relevante papel em vários processos biológicos e em várias doenças humanas. Contudo, o progresso nesta área do conhecimento tem sido limitado pela grande dificuldade em descobrir os genes que são regulados por cada miRNA, já que cada miRNA pode potencialmente ligar-se a centenas de mRNAs diferentes. Por isso, o principal objectivo desta tese consistiu em desenvolver um *framework* integrado que pudesse ser usado para descobrir os genes e vias de sinalização que são regulados por um dado miRNA. Este *framework* foi aplicado ao estudo da família miR-200, a qual já se sabia que participava no desenvolvimento embrionário e na formação de metástases, mas para a qual apenas alguns mRNAs alvo eram conhecidos.

No **Capítulo 1**, a biogénese dos miRNAs e o seu mecanismo de acção são abordados. As ferramentas computacionais e experimentais que existem para descobrir que genes são regulados pelos miRNAs são também descritas. Além disso, a transição epitelial-mesenquimal (EMT) e a via de sinalização TGF-β são também revistos pois ambos estes processos são regulados pela família miR-200. Também é feita uma revisão da família miR-200, incluindo uma lista detalhada dos genes que foram publicados como sendo regulados por esta família.

No **Capítulo 2**, o *framework* que desenvolvemos para descobrir novos genes e vias de sinalização celular regulados pelo miR-200c é descrito. Descobrímos que este miRNA potencialmente regula 520 mRNAs em ratinho. Para a maioria destes mRNAs não se sabia que eram regulados pelo miR-200c, já que no início deste projecto somente uma dezena de genes eram conhecidos como sendo regulados por este miRNA. Como esperado, estes mRNAs possuem um grande número de sequências

complementares à sequência do miR-200c e para a grande maioria deles a sua expressão é negativamente regulada pelo miR-200c. Além disso, mostrámos que este *framework* permite descobrir novos mRNAs regulados pelo miR-200c com um reduzido número de falsos positivos. Os nossos resultados mostram que o miR-200c principalmente regula 13 vias de sinalização celular. As duas vias de sinalização mais significativamente reguladas pelo miR-200c são a do receptor do factor de crescimento epidérmico (EGFR) e a da MAP cinase (MAPK). Adicionalmente, os resultados mostram que o miR-200c também regula a via de sinalização do factor beta de transformação do crescimento (TGF-β), entre outras vias. De acordo com estes resultados, mostrámos experimentalmente que o miR-200c contraria o efeito suppressor que o TGF-β e a proteína morfogenética do osso tipo 2 (BMP-2) têm na expressão de genes epiteliais. Adicionalmente, mostrámos que o miR-200c regula a região não traduzida da extremidade 3' (3'UTR) de 12 (*Crtap, Fhod1, Smad2, Map3k1, Tob1, Ywhag/14-3-3γ, Ywhab/14-3-3β, Smad5, Zfp36, Xbp1, Mapk12*, *Snail1*) de 14 mRNAs que foram escolhidos para serem validados como sendo regulados pelo miR-200c. Esta validação foi feita através da identificação precisa dos elementos que são reconhecidos pelo miR-200c no 3'UTR de cada um destes 12 mRNAs. Também mostrámos experimentalmente que o grau de interacção entre cada um destes 12 mRNAs e o miR-200c está fortemente correlacionado com o grau que cada um dos mRNAs é reprimido.

As proteínas Smad2 e Smad5 formam um complexo com a proteína Zeb2, e por sua vez as proteínas Ywhab/14-3-3β e Ywhag/14-3-3γ formam um complexo com a proteína Snail1. Estes dois complexos ligam-se aos promotores de genes epiteliais e reprimem a sua transcrição. Nós mostrámos que a sobre-expressão de miR-200c induz a ligação da RNA polimerase II aos promotores de genes epiteliais enquanto que reduz a ligação dos complexos Zeb2 e Snail1 a estes mesmos promotores.

Mostrámos também que a expressão de uma versão sintética do mRNA *Smad5* que é resistente ao miR-200c reduziu significativamente, ainda que modestamente, a indução de genes epiteliais que se verifica quando o miR-200c é sobre-expressado.

É sabido que a expressão da família miR-200 ou a supressão de *Zeb2* inibem a invasão celular em ensaios *in vitro*. Supressão de cada um dos três mRNAs que nós indentificámos como sendo regulados pelo miR-200c, *Smad5*, *Ywhag* and *Crtap*, também inibiram fortemente a invasão celular. Em conclusão, mostrámos experimentalmente que a família miR-200 suprime a via de sinalização do TGF-β/BMP, promove a expressão de genes epiteliais e inibe a invasão celular através da regulação de uma rede de genes.

No **Capítulo 3**, são discutidas as conclusões principais e as suas limitações assim como possíveis futuras linhas de investigação.

Sumarizando, nesta tese foi desenvolvido um *framework* que pode ser usado para descobrir novos genes e vias de sinalização regulados por outros miRNAs para melhor entender o seu papel biológico. É importante referir que esta tese também contribui substancialmente para o estudo da família miR-200 e da EMT pois descobriu-se novos genes e vias de sinalização que são globalmente regulados pelo miR-200c e que são usados por este para regular a EMT.

Table of Contents

CHAPTER 1: Introduction

CHAPTER 2: The role of the miR-200 family in EMT

CHAPTER 3: Discussion and conclusions

Abbreviations

Abbreviation	Full form
4SU	4-Thiouridine
Ago	Argonaute protein
Ago2	Eukaryotic translation initiation factor 2C, 2, Argonaute 2
BMP	Bone morphogenetic protein
bps	Base pairs
BSA	Bovine serum albumin
cDNA	Complimentary DNA
CDS	Coding sequence
ChIP	chromatin-immunoprecipitation
co-IP	Co-immunoprecipitation
DDX6	DEAD (Asp-Glu-Ala-Asp) box helicase 6
DMEM	Delbecco's modified essential medium
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
ERK	Extracellular regulated kinase
GTP	Guanosine 5′-triphosphate
HGF	Hepatocyte growth factor
IP	Immunoprecipitation
JNK	c-Jun N-terminal kinase
MAPK	Mitogen activated protein kinase
MET	Mesenchymal to epithelial transition
MIS	Mullerian inhibitory substance
MRE	miRNA-recognition element
miRISC	miRNA-induced silencing complex
miR/miRNA	microRNA
miRNA*	miRNA star
nt	Nucleotide

PABP	Poly(A)-binding protein
PCR	Polymerase chain reaction
PD	Pull-down
PDGF	Platelet-derived growth factor
PI3K	Phosphoinositide 3-kinase
pri-miRNA	Primary miRNA
qRT-PCR	Quantitative real-time PCR
RhoB	Ras homolog gene family, member B
RNA	Ribonucleic acid
RNAse	Ribonuclease
shRNA	Short-hairpin RNA
siRNA	Small-interfering RNA
S1P	Sphingosine-1-phosphate
Smad	Small mothers against decapentaplegic
Smurf	SMAD-ubiquitination-regulatory factor
snoRNA	Small-nucleolar RNA
TALEN	Transcription activator-like effector nucleases
TGF-β	Transforming growth factor-beta
tRNA	Transfer RNA
UTR	Untranslated region
Zeb1	Zinc finger E-box binding homeobox 1
Zeb2	Zinc finger E-box binding homeobox 2

List of Figures

List of Tables

Chapter 1

Introduction

CONTENTS

1. microRNAs

1.1. Function

microRNAs (miRNAs) are short (~22-nucleotide (nt)-long), noncoding single-stranded RNA molecules that regulate gene expression by binding to partially complementary sequences in target mRNAs to induce mRNA decay, translational repression or both[1]. Each miRNA can potentially bind to hundreds of mRNAs, sometimes having a strong effect but more often having a weak fine-tuning effect. miRNAs participate in a variety of biological processes such as development, differentiation, proliferation, antiviral defense, metabolism and cellular stress signaling[1-2]. At least 60% of human protein-coding genes are regulated by miRNAs (mostly downregulated). Dysregulation of miRNA expression levels has been associated with several pathologies such as cancer, neurodevelopmental, autoimmune, liver, muscle, skin and cardiovascular diseases[3-4]. Also, mutations in miRNA genes have been shown to cause genetic diseases such as nonsyndromic autosomal dominant progressive hearing loss which is induced by point mutations in the seed sequence of miR-96. Likewise, mutations in miR-196a-2 elevate the risk of developing lung, breast, liver and gastric cancer.

1.2. Biogenesis and action mechanism

miRNAs are first transcribed from the genome by RNA polymerase II into capped and polyadenylated primary miRNA (pri-miRNA) transcripts[5] (Figure 1).

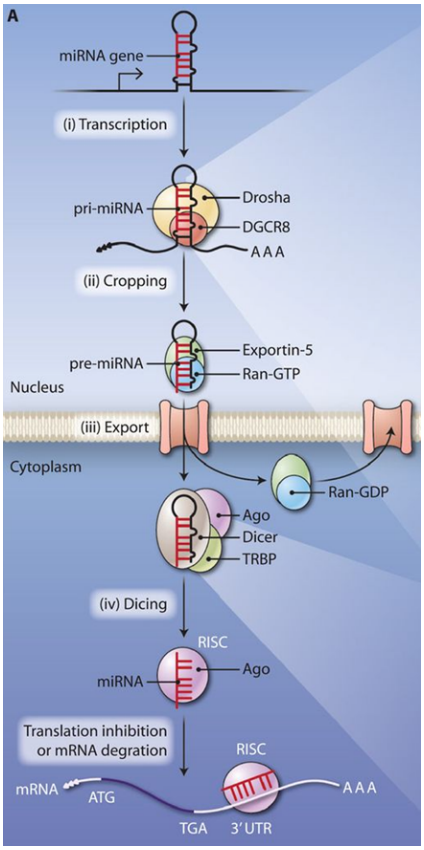


Figure 1. Mechanisms that regulate miRNA biogenesis and RISC loading. (A) Biogenesis of a miRNA is a stepwise process involving (i) transcription of a primary transcript (pri-miRNA), (ii) processing ("cropping") by the Drosha/DGCR8 microprocessor complex to produce a pre-miRNA, (iii) export from the nucleus into the cytoplasm, mediated by the exportin XPO5 and the guanosine 5′-triphosphate (GTP)–binding protein Ran, and (iv) additional processing ("dicing") by a complex containing the RNase Dicer, a catalytic component, Argonaute (Ago), and the RNA-binding protein TRBP (transactivation response RNA binding protein) to produce the mature miRNA. The resulting complex of Ago and the mature, single-stranded miRNA is called the RISC. The miRNA binds complementary sequences in mRNAs; Ago either inhibits the translation of the mRNA or cleaves it to cause its degradation. Figure and legend reproduced from Hata A, Lieberman J. *Sci Signal* 2015[6].

Intragenic miRNAs (located within introns, exons or untranslated regions (UTRs) of protein-coding transcripts) are transcribed with their host genes through shared common promoters[7]. Alternatively, intergenic miRNAs are

transcribed from individual, non-protein coding genes which have their own promoters[8].

These pri-miRNA transcripts contain one or more stem-loop structures (also called hairpin structures) in which miRNAs are embedded[2]. Still in the nucleus, these stem-loop structures are recognized by the double-stranded RNA-binding protein DGCR8 (DiGeorge syndrome critical region 8), also known as Pasha in invertebrates. DGCR8 associates with Drosha, a ribonuclease (RNase) III family enzyme, and together they form the microprocessor complex. Drosha then cleaves the pri-miRNA into a precursor miRNA (pre-miRNA). The pre-miRNA (~70-nt long), is then transported to the cytoplasm by a nuclear complex formed by Exportin-5 together with the guanosine 5′-triphosphate (GTP)–binding protein Ran. Once in the cytoplasm, the pre-miRNA is cleaved by Dicer (another RNase III enzyme) together with the RNA-binding protein TRBP (transactivation-response RNA-binding protein) to generate an RNA duplex. This duplex consists of the mature miRNA (depicted in red in Figure 1), and the miRNA star (miRNA*, RNA strand partially complementary to the mature miRNA). While the mature miRNA is loaded into an Argonaute (Ago) protein, the miRNA* is usually released and degraded. Ago proteins also bind to members of the GW182 protein family (in humans there are three GW182 paralogs known as TNRC6A-C) and this complex bound to a miRNA forms the miRNA-induced silencing complex (miRISC)[6]. Ago proteins are the core components of the miRISC complex as they bind to the miRNA and mediate recognition of its target mRNAs.

In addition to this canonical miRNA biogenesis pathway some alternative pathways have been discovered. Some miRNAs are generated by Drosha- or Dicer independent mechanisms, such as mirtrons (through the splicing of miRNA-containing introns) and miRNAs originated from small-nucleolar RNAs (snoRNAs) and transfer RNAs (tRNAs)[9].

The mature miRNA then guides the miRISC complex to partially complementary sequences in target mRNAs (mostly in 3'UTRs) termed miRNA-recognition elements (MREs). Base pairing between miRNAs and their target mRNAs is most critical within the miRNA seed region (nucleotides 2-8), but additional base pairing beyond this region also occurs. If base pairing between a miRNA and its target mRNA is extensive (within and outside the seed region) then the target mRNA is cleaved through Ago-catalyzed mRNA cleavage[10-11]. This type of repression is dominant in plants[12], but in animals the majority of targets lack the extensive base pairing required for Ago-mediated cleavage[13] (in animals this infrequent Ago-mediated mRNA cleavage is catalyzed by AGO2, the most abundant of the four Ago family members[14]). In animals miRISC complexes more frequently silence mRNAs through a combination of mRNA translational repression and mRNA degradation, with the latter contributing to most of the repression (66-90%)[15]. The precise molecular mechanism of miRISC-induced translational repression is still not completely understood, but the current model is that miRISC induces mRNA translational repression by interfering with the activity and/or assembly of the eukaryotic translation initiation factor 4F (eIF4F) complex that is required for cap-dependent mRNA translation. The eIF4F complex consists of the cap-binding protein eIF4E, the adaptor protein eIF4G and the DEAD (Asp-Glu-Ala-Asp) box RNA helicase eIF4A[16]. Alternatively or in combination with this process of mRNA translational repression, miRISC can induce mRNA target degradation. Target mRNA degradation is achieved through recruitment of GW182 proteins by miRNA-loaded Ago proteins. GW182 proteins in turn interact with the cytoplasmic poly(A)-binding protein (PABPC) and with the cytoplasmic deadenylase complexes PAN2-PAN3 and CCR4-NOT, which catalyze the deadenylation of target mRNAs. Deadenylated mRNAs are then decapped and rapidly degraded by the 5'-to-3' exoribonuclease 1 (XRN1).

1.3. Computational prediction of miRNA targets

It is now established that miRNAs have a critical role in virtually all biological processes and that at least 60% of all genes are regulated by miRNAs[4]. Therefore, it is extremely important to predict the biologically relevant miRNA–mRNA interactions in any given biological context. For this end, several computational prediction tools have been developed. Computational methods exploit miRNA-targeting features to provide a genome-wide prediction of miRNA targets. In animals, as opposed to plants, miRNAs pair imperfectly to almost all target mRNAs and therefore computational prediction of miRNA targets is more complex. Commonly used computational methods (Table 1) place variable weight on base pairing between the miRNA-seed region and target mRNA[17], evolutionary conservation of the MRE[4,18], thermodynamic stability of the miRNA/mRNA heteroduplex[19] and mRNA sequence features outside the MRE[20]. Computational miRNA-target prediction methods are valuable for the preliminary identification of miRNA targets but there are inherent limitations to consider when applying them to discover new miRNA targets. The main limitation is their high false-positive rates, as even the best algorithms such as TargetScan typically predict thousands of target genes for any given miRNA. The false-positive rates of target prediction tools such as TargetScan, PicTar and miRanda are between 22 and 31%[21]. These high false-positive rates are mainly due to the fact that computational target-prediction tools use sequence determinants to predict miRNA targets. However, it has been shown that sequence features alone cannot provide accurate prediction of miRNA targets as experimental analyses of Ago-bound miRNA-mRNA heteroduplexes suggest that 25%–45% of MREs lack a perfect seed match[22-23] and a growing number of experimentally-validated MREs lack a canonical seed[24-27]. Also, most computational tools limit target prediction to the mRNA 3′UTR even though there is experimental evidence

suggesting that as many as 25% of functional MREs are in mRNA coding regions[23].

Table 1. Characteristic features of algorithms providing information beyond classical miRNA target gene interaction. Table reproduced from Tarang S, Weston MD *RNA Biology* 2014[28].

Prediction method	Characteristic features	Resource
Diana Micro-T	Target prediction made with miRNA or mRNA sequences as input	diana.cslab.ece.ntua.gr/microT/
FAME	Experimentally verified miRNA pathways infer biological process affected by miRNAs	acgt.cs.tau.ac.il/fame/
Hoctar	Information on host genes regulating expression of its embedded microRNA	hoctar.tigem.it
Magia	Query of miRNA target prediction, analysis of expression profiles, and post-transcription regulatory network	gencomp.bio.unipd.it/magia/start/
MaMi	miRNA target prediction based on hybridization energies and secondary structures for the miRNA-mRNA hybrid where parameters can be modified by the user	mami.med.harvard.edu
Microinspector	Identification of potential miRNA binding sites in user-submitted sequences, searching against databases of known miRNA binding sites	bioinfo.uni-plovdiv.bg/microinspector/
miR2disease	Information on miRNA association to a disease process	www.mir2disease.org
miRBase	Complete repository of miRNA sequences and targets	www.mirbase.org
mirDIP	mirDIP integrates 12 microRNA prediction data sets from six microRNA prediction databases	ophid.utoronto.ca/mirDIP/index.jsp
miRecords	Two main modules, experimentally validated targets, and integrated information across 11 independent prediction softwares	mirecords.umn.edu/miRecords/index.php
miRGator	Integrates miRNA expression data with mRNA and protein to interpret the biological functions of miRNAs	genome.ewha.ac.kr/miRGator/miRGator.html
miRNAmap	Information on statistics of miRNA sequences and target genes	mirnamap.mbc.nctu.edu.tw
MirPath	Identification of altered molecular pathways by the expression of specific miRNAs	83.212.96.7/DianaToolsNew/index.php?r=mirpath
miRTar	Information on the biological function of miRNA-target gene pairs and on miRNA sites on the alternative spliced transcripts	mirtar.mbc.nctu.edu.tw/human/
MiRTarBase	Information on experimentally verified miRNA targets by data mining and manually surveying pertinent literature related to functional studies on miRNAs	mirtarbase.mbc.nctu.edu.tw
miRWalk	miRNA-target information on the complete sequence (promoter, 5′-UTR, CDS, and 3′-UTR) and target interaction information across eight other types of prediction software; information on experimentally validated targets	www.umm.uni-heidelberg.de/apps/zmf/mirwalk/index.html
Patrocles	Polymorphisms on miRNA sequences and target genes	www.patrocles.org
PITA	Secondary structure of the miRNA-mRNA hybrid for target gene prediction	genie.weizmann.ac.il/pubs/mir07/index.html
Tarbase	Experimentally validated miRNA targets	diana.cslab.ece.ntua.gr/tarbase/
Targetscan	Classical software for miRNA target gene prediction	www.targetscan.org

Also, computational analyses need to take into account cell-specific mRNA isoforms. In biological processes such as proliferation[29], differentiation[30] and cancer[31] some mRNAs are alternatively polyadenylated and generate transcripts with unique 3′UTRs, which may have different MREs. Therefore, high-throughput mRNA sequencing of the biological system of interest combined with computational prediction of miRNA targets can more accurately identify functional MREs[32].

1.3.2. Experimental identification of miRNA targets

Experimental identification and validation of miRNA targets is critical to identify miRNA biological functions accurately. This is the most challenging and important issue in this field as miRNA functions are determined by their target genes. During the past decade several strategies were developed to improve experimental miRNA target identification but still relatively few targets have been thoroughly experimentally validated. One of the reasons is that miRNA targeting is often context-, time-, or tissue-dependent and therefore some mRNAs might only be targeted in some contexts even though they have extensive sequence complementarity with a given miRNA. Some experimental approaches used to identify miRNA targets will be briefly described below (Figure 2).

1.3.2.1. Reporter assays

One common approach to test if a given mRNA can be targeted by a specific miRNA is to use a dual-luciferase reporter assay[33]. In this method, a putative mRNA target sequence (e.g. full or partial 3'UTR) is inserted downstream of the luciferase coding sequence before the polyadenylation signal. In order to normalize the luciferase luminescence level, two luciferase genes with different light-emission wavelengths are usually used, a constitutively-expressed Firefly luciferase and a *Renilla* luciferase containing the putative target sequence. The ratio of *Renilla* to firefly-

luciferase activity level is used to determine the extent of miRNA target repression after cells are transfected with both luciferase reporter genes together with the miRNA of interest.

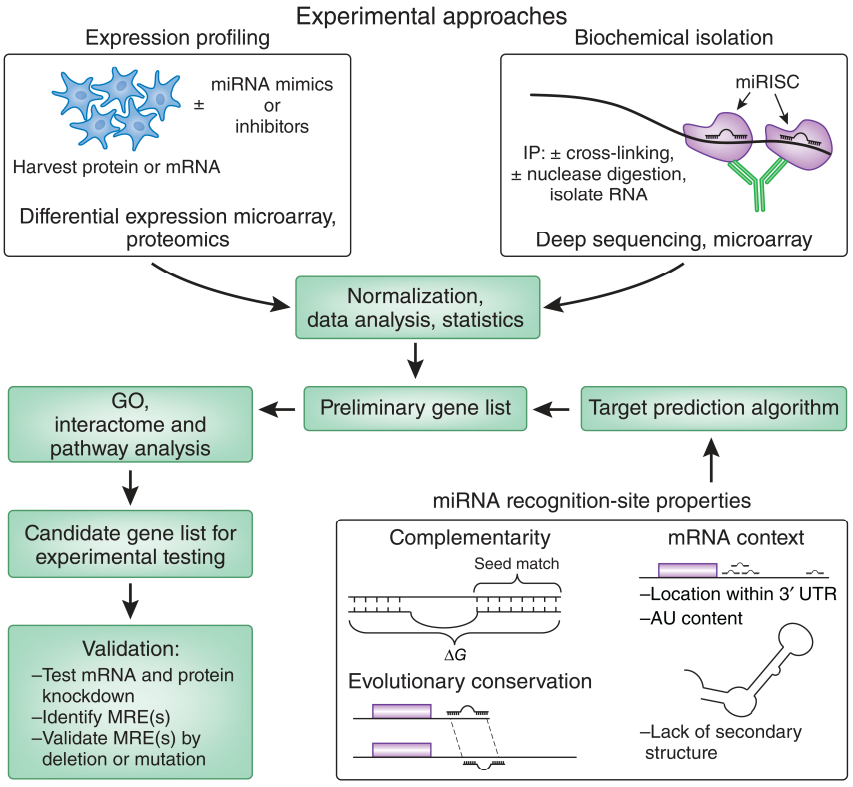


Figure 2. Methods for identifying miRNA targets. Putative target genes can be identified by expression profiling of cells in which the miRNA is overexpressed or antagonized, by biochemical isolation of the miRISC or by target prediction algorithms. These methods generally identify hundreds of candidate genes or more. Bioinformatic analysis of these large candidate gene lists for over-represented Gene Ontology (GO) terms, enriched biological pathways or gene interaction networks can then help researchers to select candidate genes to evaluate experimentally. Figure and legend reproduced from Thomas M, Lieberman J, Lal A. Nat Struct Mol Biol 2010[1].

As a negative control, similar constructs in which the putative MRE was mutated can be used. Even though a dual-luciferase reporter assay should always be used to confirm direct miRNA targeting and identify functional MREs it does not definitely show that a given mRNA is targeted by a specific miRNA *in vivo*, as the supraphysiological miRNA concentrations

used and the inherent limitations of reporter assays might result in artifacts. Also, this method is low-throughput, expensive and labor intensive and therefore not appropriate to identify miRNA targets globally.

1.3.2.2. mRNA and protein expression profiling

One common approach to identify genome-wide miRNA targets is to measure global mRNA/protein levels after overexpressing or inhibiting a given miRNA, as its targets' mRNA/protein often change (Figure 2). However, measuring global mRNA/protein levels after manipulating a given miRNA doesn't allow to distinguish between direct and indirect targets as changes at the mRNA or protein level may be due to downstream effects instead of direct miRNA targeting.

Global mRNA levels can be measured by microarray or more recent high-throughput mRNA sequencing. To measure global miRNA effects at the protein level, proteins can be labeled using a stable isotope (stable-isotope labeling with amino acids in cell culture, SILAC) followed by mass spectrometric analysis[34]. However, this approach is time-consuming, expensive and only a fraction of the proteome can be resolved at a time. Therefore, microarray analysis is often used instead, as mRNA and protein levels are in most cases correlated.

There are several approaches to manipulate intracellular miRNA concentration. miRNAs can be overexpressed by transfecting cells with synthetic miRNA duplexes that contain the miRNA of interest. Once the synthetic miRNA duplex is assembled into the miRISC it behaves as the endogenous miRNA. Alternatively, miRNAs can be antagonized for loss-of-function studies using several strategies. First, for permanent and complete miRNA knockdown, transcription activator-like effector nucleases (TALENs) or CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat/ CRISPR associated protein 9) technology can be used to specifically edit

miRNA genes and totally abrogate miRNA expression[35]. Second, antagomirs, which are chemically modified single-stranded RNA oligonucleotides complementary to target miRNAs, are able to bind and prevent miRNAs from interacting with their targets, thereby upregulating their mRNA levels[36-37]. In addition to sequester miRNAs, antagomirs can instead bind to a specific target mRNA sequence and prevent miRNAs from binding to it, which also upregulates target mRNA levels[38]. Lastly, miRNAs can be inhibited by using miRNA sponges. miRNA sponges are transcripts that contain multiple, tandem binding sites complementary to a miRNA of interest[39]. When plasmid vectors encoding miRNA sponges are transfected into cells, sponges can derepress miRNA target mRNAs, in some cases as strongly as antagomirs. It is important to keep in mind that all these miRNA manipulations have their caveats. Transfecting excessive amounts of synthetic miRNA duplexes can be toxic to cells or may introduce off-target effects[40-41]. Also, miRNA overexpression above physiological levels may lead to artifacts, as excess miRNA can saturate all available miRISC in the cell and affect the endogenous miRNA pathway[40]. On the other hand, inhibition of a highly expressed miRNA might be difficult to achieve with antagomirs and miRNA sponges, and might only be achieved with more laborious knockout technology. Also, inhibition of a poorly expressed miRNA might not have a detectable effect on its target genes.

1.3.2.3. Immunoprecipitation of miRISC components

Because of the direct interaction between Ago proteins, miRNAs and their target mRNAs in the miRISC complex, direct targets can be identified by immunoprecipitating individual miRISC components[42-43], such as Ago or Tnrc6 protein-family cofactors[44], followed by identification of the pulled-down mRNAs by microarray or high-throughput RNA sequencing. For these studies the miRNA of interest is usually overexpressed to saturate all available miRISC and isolate specific miRNA targets. While some studies

perform immunoprecipitation (IP) of endogenous miRISC components[42] others overexpress epitope-tagged miRISC components inside cells before immunoprecipitating[43,45-47]. All these studies have high false-positive rates (~40-70%[23]) for several reasons. First, target mRNAs identified by immunoprecipitating a specific Ago family member might not be identical to those bound by other Ago family members[48] (in humans there are four Ago-family members (Ago1-4)[49]). Second, endogenous and epitope-tagged Ago proteins bind to transfer RNAs[50] and overexpression of Ago proteins has been shown to increase global miRNA production,[51-52] which might result in artifacts. Additionally, as Ago proteins can still bind to mRNAs *in vitro* in cell extracts after cell lysis[46] some of the mRNAs isolated through an Ago IP might not be targets *in vivo*. Moreover, Ago IPs only allow identification of putative target mRNAs without providing any information on the precise location of functional MREs in those mRNAs. Also, Ago IPs are usually not completely specific for the miRNA of interest, even after overexpression of the miRNA, as there are always other endogenous miRNAs still bound to miRISC complexes that may skew results. Finally, Ago IPs require miRNAs to be strongly bound to their target mRNAs and therefore this assay has low sensitivity for weak or transient miRNA/mRNA interactions.

To overcome these limitations a recently-developed technique called Ago HITS-CLIP (high-throughput sequencing by crosslinking and immunoprecipitation) can be used instead to identify functional MREs[22-23]. In this technique, short-wavelength (254 nm) UV light is firstly used to crosslink miRISC proteins to any bound RNA. RNA is then isolated by immunoprecipitating one of the miRISC components. RNA in the immunoprecipitate undergoes partial digestion to remove RNA that is not miRISC-protected. Then RNA fragments that are crosslinked to the miRISC go through high-throughput RNA sequencing to identify both Ago-associated miRNAs and their target MREs. This method identifies direct miRNA targets with low false-positive rates (~13–27%[23]). However, the use

of short-wavelength UV light has low crosslinking efficiency, can produce non-specific crosslinks and may photo-damage RNAs.

A new approach called PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) has been developed to overcome the limitations of previous CLIP methods[53]. In this technique (Figure 3), a photoreactive ribonucleoside analog (such as 4-thiouridine (4SU), which can substitute for uridine during transcription) is supplied to cells and incorporated into nascent transcripts. This modified nucleoside facilitates crosslinking by UV radiation at 365 nm and increases RNA yields compared to the conventional short-wavelength (254 nm) UV light crosslinking without 4-thiouridine[53]. The 4SU nucleoside analog allows precise identification of functional MREs as 4SU preferentially binds to guanine instead of adenine during the cDNA generation step. This results in thymine to cytosine (T-to-C) mutations after the PCR-amplification step (Figure 3). These T-to-C mutations will be detected during sequencing by comparing the sequencing reads to a reference genome. These T-to-C mutations are then used to identify which mRNA sequences (MREs) were bound by Ago *in vivo*.

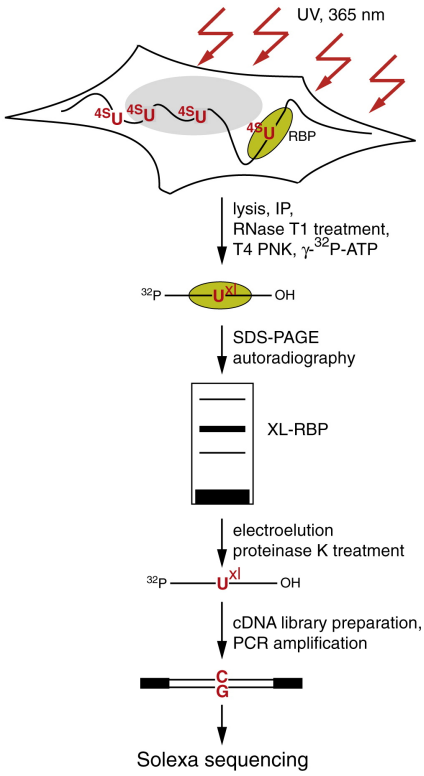


Figure 3. Illustration of PAR-CLIP. 4SU-labeled transcripts were crosslinked to RBPs and partially RNase-digested RNA-protein complexes were immunopurified and size-fractionated. RNA molecules were recovered and converted into a cDNA library and deep sequenced. Figure and legend reproduced from Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P *et al*. Cell 2010[53].

Recently, a modified CLIP method, iCLIP (individual-nucleotide resolution CLIP), was introduced to increase resolution of MRE identification[54] (Figure 4). This method does not use photo-reactive nucleosides and their associated T-to-C mutations to determine miRNA binding sites. Instead, it takes advantage of the fact that some cDNAs are truncated at the UV cross-linking site as Proteinase K digestion leaves a covalently bound polypeptide fragment on the RNA that causes premature truncation of the reverse transcription reaction at the cross-link site. Therefore, cDNA circularization is used to reveal the exact nucleotide at the crosslinking site via high-throughput sequencing.

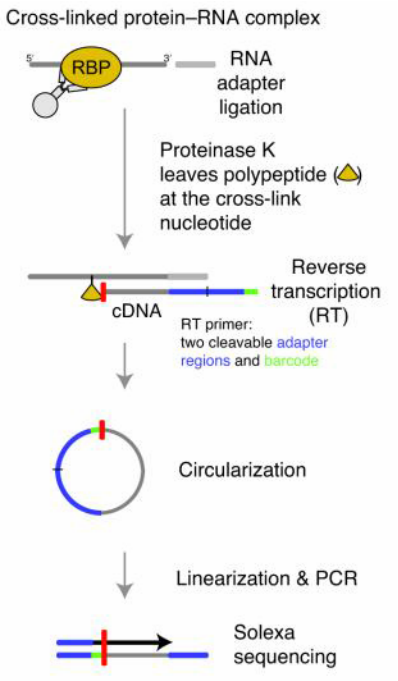


Figure 4. Schematic representation of the iCLIP protocol. After UV irradiation, the covalently linked RNA is co-immunoprecipitated with the RNA-binding protein (RBP) and ligated to an RNA adapter at the 3′ end. Proteinase K digestion leaves a covalently bound polypeptide fragment on the RNA that causes premature truncation of reverse transcription (RT) at the cross-link site. The red bar indicates the last nucleotide added during reverse transcription. Resulting cDNA molecules are circularized, linearized, PCR-amplified and subjected to high-throughput sequencing. The first nucleotides of each sequence contain the barcode followed by the nucleotide where cDNAs truncated during reverse transcription. Figure and legend reproduced from Sugimoto Y, Konig J, Hussain S, Zupan B, Curk T, Frye M *et al*. Genome Biol 2012[54].

For each of these different CLIP methods laborious bioinformatic analyses have to be performed to finally miRNA targets, as these CLIP methods are not able to identify which specific miRNA was associated with each Ago-bound RNA that was immunoprecipitated. Extensive bioinformatic analyses are then performed to the sequencing data to assign a miRNA to each immunoprecipitated RNA fragment based on the assumption that each of these RNA fragments has to have a sequence complementary to the seed sequence of the miRNA that was bound to it. Since many MREs do not have an exact seed pairing sequence, this bioinformatics method discards many *bona fide* miRNA/mRNA interactions. Therefore, non-specific

Ago/RNA interactions can result in false-positives. Also, these CLIP methods are technically-challenging multistep protocols.

1.3.2.4. Pull-down of biotinylated miRNAs

To address these problems, recent studies[55-60] have used a much simpler approach that avoids cross-linking that consists on transfecting cells with a synthetic, biotinylated version of the miRNA of interest followed by streptavidin pull-down (PD) of its target RNAs (Figure 5). The synthetic miRNA mimics (RNA duplexes consisting of the mature miRNA of interest and its partially complementary passenger strand) have a biotin molecule covalently attached to the 3′ end of the mature miRNA strand (also called antisense strand). The pulled-down RNA can then be identified by microarray or high-throughput RNA sequencing. Putative miRNA targets can then be easily identified by comparing the enrichment level of each identified RNA in the PD performed with the miRNA of interest versus a control PD performed with an unrelated miRNA without any predicted targets in the system of interest (for example, the unrelated *C. elegans* miR-67 is usually used as a control miRNA mimic for PDs performed in mammalian cells). Even though the transfected synthetic-miRNA mimic will be present at much higher intracellular levels than the corresponding endogenous miRNA, recent evidence suggests that this still allows identification of physiologically-relevant miRNA/mRNA interactions. In fact a 10-fold increase in miRNA concentration produces only a 10% increase in target repression[61]. This simple and fast method has been used to identify miRNA targets with high specificity[55-60].

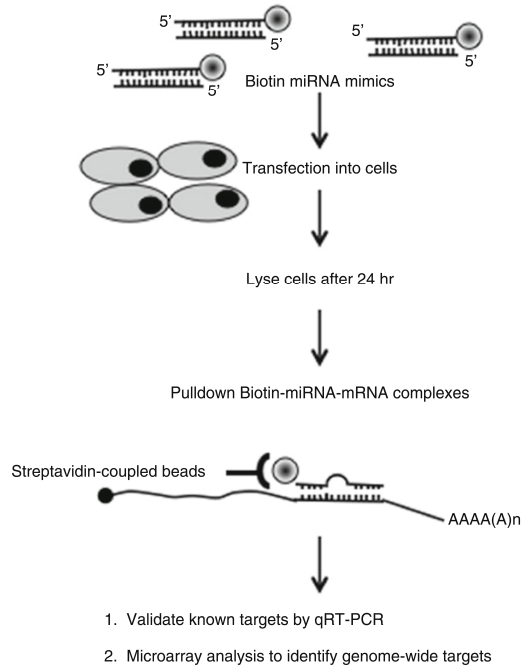


Figure 5. Overview of the pulldown method. 3′-Biotinylated miRNA mimics are transfected into cells and cytoplasmic extracts are prepared after 24h. The Biotin–miRNA–mRNA complexes are isolated from cytoplasmic extracts using streptavidin beads. RNA isolated from the pulldown material is used to determine the enrichment of known targets of the miRNA. Microarray analysis from the pulldown RNA can be used to identify the genome-wide mRNAs bound by the transfected Biotin-miRNA. Figure and legend reproduced from Subramanian M, Li XL, Hara T, Lal A. Methods Mol Biol 2015[62].

2. Epithelial to mesenchymal transition

Epithelial to mesenchymal transition is a complex process that converts polarized epithelial cells into migratory and invasive mesenchymal cells[63] (Figure 6). EMT is a normal process during embryo development and is required for the formation of the neural crest[64], fusion of the hard palate[65], and other developmental processes. However, in tumorigenesis EMT is a pathological event that allows carcinomas to metastasize[66-68].

EMT is a complex process at the molecular level, with multiple biological pathways involved from its initiation until its completion. These pathways

comprise activation of several transcription factors, expression of specific cell surface proteins, expression and reorganization of cytoskeletal proteins, production of metalloproteinases that degrade extracellular matrix (ECM) components, as well as changes in miRNA expression profiles[63].

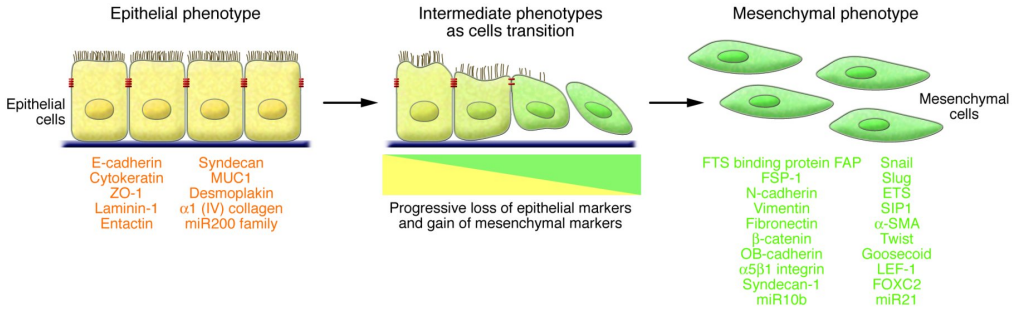


Figure 6. Epithelial to mesenchymal transition (EMT). An EMT involves a functional transition of polarized epithelial cells into mobile and ECM component-secreting mesenchymal cells. The epithelial and mesenchymal cell markers commonly used by EMT researchers are listed. ZO-1, zona occludens 1; MUC1, mucin 1, cell surface associated; miR200, microRNA 200; SIP1, survival of motor neuron protein interacting protein 1; FOXC2, forkhead box C2. Figure and legend reproduced from Kalluri R, Weinberg RA. J Clin Invest 2009[63].

EMT is activated under normal or pathological biological contexts and therefore some authors[63,69] classify it into three different subsets. Type 1 EMT occurs during the formation and implantation of the embryo and organ development. Type 2 EMT is associated with wound healing, tissue regeneration and fibrosis. This type of EMT is a repair mechanism that normally generates fibroblasts to replace and restore tissue integrity after inflammation or injury. In this case EMT ceases when the inflammation is mitigated in the case of wound healing and tissue regeneration. Type 3 EMT is a pathological event that contributes to the ability of carcinomas to metastasize (Figure 7). About 90% of cancer-related deaths are caused by metastasis. For example, breast cancer begins as a local disease and later metastasizes to lymph nodes and vital organs such as lung, liver, bone and

brain[70]. Metastasis is a complex, multistep process that starts when epithelial cells in the primary tumor undergo EMT and acquire a mesenchymal phenotype. These cells are then able to detach from their neighboring cells, such as nearby stromal cells, and invade through the basement membrane (Figure 7). Subsequently, metastasizing cells enter the bloodstream or the lymphatic system (intravasate) and then exit (extravasate) at distant organs forming micrometastases. The reversal of the EMT process, mesenchymal to epithelial transition (MET), is also an important event in cancer progression, as it has been hypothesized that MET is required for the final step of metastasis formation, the colonization of the metastatic site by proliferation of the established micrometastases[71-72]. Although great advances have been made in cancer treatment, particularly in its early stages, metastasis remains a challenging and frequently fatal process[73]. Therefore, blocking the metastatic process holds great therapeutic promise.

Many cytokines induce EMT, including hepatocyte growth factor (HGF), epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factor-beta (TGF-β) and bone morphogenetic proteins (BMPs). These cytokines activate transcription factors such as Zeb1 (zinc finger E-box binding homeobox 1), Zeb2 (zinc finger E-box binding homeobox 2, also known as SIP1), Twist (Twist1), Snail (Snail1), Slug (Snail2), Hmga2 (high mobility group A2) and Sp1 (specificity protein 1)[68,74-75]. These transcription factors activate downstream signaling molecules and pathways including mitogen activated protein kinases (MAPKs), phosphoinositide 3-kinase (PI3K), Akt (also known as protein kinase B), small mothers against decapentaplegic (Smad) protein family, RhoB (Ras homolog gene family, member B), β-catenin, Ras, c-Fos and integrin signaling[63]. Several of these factors transcriptionally repress E-cadherin, an important cell adhesion molecule. Loss of E-cadherin is an hallmark of oncogenic EMT and knockdown of E-cadherin alone is sufficient to induce

EMT and metastasis[76], highlighting the importance of E-cadherin regulation during EMT.

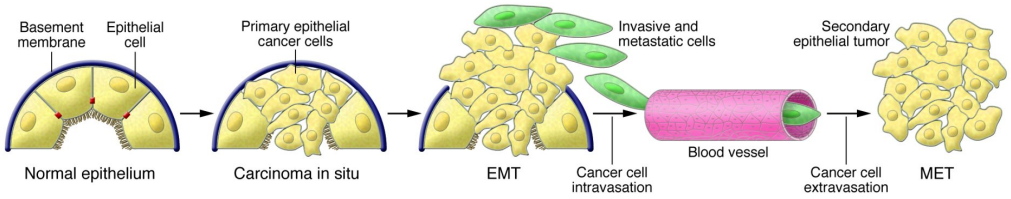


Figure 7. EMT contribution to cancer progression. Progression from normal epithelium to invasive carcinoma goes through several stages. The invasive carcinoma stage involves epithelial cells losing their polarity and detaching from the basement membrane. The composition of the basement membrane also changes, altering cell-ECM interactions and signaling networks. The next step involves EMT and an angiogenic switch, facilitating the malignant phase of tumor growth. Progression from this stage to metastatic cancer also involves EMTs, enabling cancer cells to enter the circulation and exit the blood stream at a remote site, where they may form micro- and macro-metastases, which may involve METs and thus a reversion to an epithelial phenotype. Figure and legend reproduced from Kalluri R, Weinberg RA. J Clin Invest 2009[63].

2.1. TGF-β superfamily signaling

The transforming growth factor-β (TGF-β) protein superfamily has an important role in development and carcinogenesis. The TGF-β protein superfamily consists of four main subgroups of structurally-related protein families that share a common signaling pathway. The TGF-β protein superfamily includes the TGF-β family, the bone morphogenetic protein (BMP) family, the inhibin/activin family and the mullerian inhibitory substance (MIS) family[77]. Mechanistically, this signaling pathway is activated when a TGF-β superfamily ligand (TGF-β, BMPs, etc) binds to a TGF-β superfamily type II receptor. Ligand-bound type II receptors become activated and phosphorylate the cytoplasmic domain of type I receptors, consequently activating them (Figure 8). Vertebrates have seven distinct type I receptors, each of which can assemble with one of five type II receptors to mediate signaling for each specific TGF-β superfamily ligand.

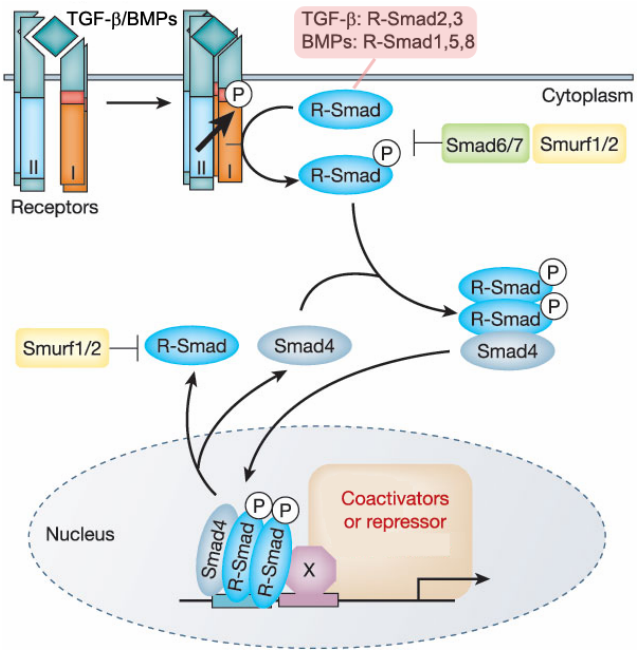


Figure 8. General mechanism of TGF-β superfamily signaling and Smad activation. At the cell surface, a TGF-β superfamily ligand binds a complex of transmembrane receptor serine/threonine kinases (types I and II) and induces transphosphorylation of the type I receptor by the type II receptor kinases. The consequently activated type I receptors phosphorylate selected Smads (Smad2 or -3 for TGF-β and Smad1, -5 or -8 for BMPs), and these receptor-activated Smads (R-Smads) then form a complex with a common Smad4. Activated Smad complexes translocate into the nucleus, where they regulate transcription of target genes, through physical interaction and functional cooperation with DNA-binding transcription factors (X) and CBP or p300 coactivators. Activation of R-Smads by type I receptor kinases is inhibited by Smad6 or Smad7. R-Smads and Smad4 shuttle between nucleus and cytoplasm. The E3 ubiquitin ligases Smurf1 and Smurf2 mediate ubiquitination and consequent degradation of R-Smads. Figure and legend adapted from Derynck R, Zhang YE. Nature 2003[78].

The activation of type I receptors leads to Smad protein phosphorylation and activation. These receptor-phosphorylated Smad proteins are also called receptor regulated Smads or R-Smads. For example, TGF-β induces phosphorylation of Smad2 or -3 while BMPs activate Smad1, -5 or -8. These activated Smad proteins then form a complex with Smad4. Together, this complex translocates into the nucleus where it transcriptionally activates or represses target genes by binding to several co-factors[79]. Many EMT-promoting transcription factors can act as Smad cofactors to repress

or activate EMT-related genes (Table 2). Of note, other intracellular signaling pathways besides the Smad signaling pathway respond to TGF-β superfamily ligands such as the MAPK, Erk and JNK pathways[78].

Table 2. EMT-promoting Smad complexes and their target genes

Cofactor	Smad partner	Target genes	Transcriptional activity type	Ref.
Snail1	Smad3 or -4	E-cadherin, occludin, claudin-3, claudin-4	Repression	[80]
Zeb1	Smad2, -3 or -4	E-cadherin, claudin-3, claudin-4	Repression	[81]
Zeb2	Smad1, -2, -3 or -5	E-cadherin, claudin-3, claudin-4	Repression	[82]
Sp1	Smad3	Endoglin, collagen, vimentin	Activation	[74]
Hmga2	Smad2 or -4	Snail1	Activation	[75]

The TGF-β superfamily signaling is inhibited by molecules such as the Sloan-Kettering Institute proto-oncogene (Ski), the Ski-related novel gene, non-Alu-containing (SnoN) and the transforming growth interacting factor (TGIF) which have been shown to prevent gene transcription through inhibition of R-Smads[83-84]. Also, two other Smad protein family members, Smad6 and Smad7, are called inhibitory Smads as they antagonize R-Smad phosphorylation by blocking their access to type I receptors and/or promote the degradation of the receptor complex[85-86]. Additionally, TGF-β superfamily signaling is inhibited by Smurf1 (Smad-ubiquitination-regulatory factor 1) and Smurf2 which interact with R-Smads and target them for degradation[87].

3. miR-200 family

The miR-200 family is composed by five homologous miRNAs (miR-141, miR-429, miR-200a, miR-200b and miR-200c), whose sequences are completely conserved between humans and mice. These miRNAs belong

to two subfamilies, whose seed regions (nucleotides 2 to 7 of the mature active strand, which is the most important region for miRNA target recognition), differ by a single nucleotide (Figure 9).

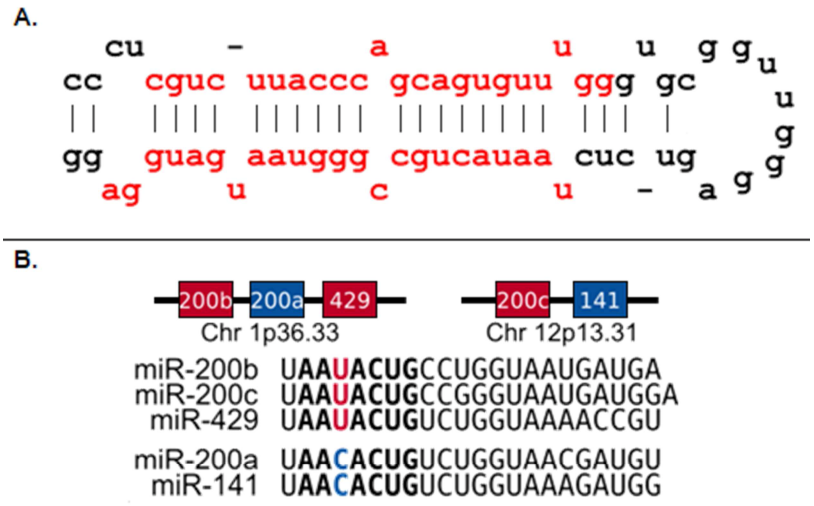


Figure 9. miR-200 family. **A**. Stem-loop structure of hsa-mir-200c (precursor miRNA). **B**. The miR-200 family members. The human miR-200 family is located in two fragile chromosomal regions on 1p36.33 (200b, 200a and 429) and 12p13.31 (200c and 141), respectively. It consists of two clusters based on seed sequence similarity: miR-200bc/429 (red) and 200a/141 (blue), distinguished by a single nucleotide change (U to C). Figure and legend reproduced from Mutlu M SÖ, Raza U, Eyüpoglu E, Yurdusev E, Sahin Ö. Atlas Genet Cytogenet Oncol Haematol 2015[88].

The seed for miR-200b/c/429 sequence is 5'-AA**U**ACU-3', while for miR-200a/141 is 5'-AA**C**ACU-3'. Additionally, they are encoded in two different gene clusters in humans and mice (in humans miR-200c and miR-141 are located on chromosome 12, while miR-200b, miR-200a and miR-429 are on chromosome 1; in mice, miR-200c and miR-141 are located on chromosome 6, while miR-200b, miR-200a and miR-429 are on chromosome 4).

3.1. Published miR-200 family target genes

During the progression of this project various miR-200 family target genes have been described in humans and mice (Table 3) in addition to those identified in this project[89] (see Chapter 2).

Table 3. Published miR-200 target genes. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown.

	Gene	Species	miR-200 family member	Ref.
1	ACVR2B	H	200c, 141	90
2	AP-2A (TCFAP2A)	H	200c, 200b, 429	91
3	ARHGAP19	H	200c	92
4	BCL-2	H	200b, 200c, 429	93
		H	429	94
5	BMI1	H	200b	95-96
		H	200c	97-99
		H	141	100
		M	200c	101
6	BRD3	H	141	102
7	BRD7	H	200c	103
8	CAGE	H	200b	104
9	CD47	H	141	105
10	CDC25B	H	141	106
11	CDH11	H	200c	107
		M	141	108
12	CDK2	H	200b	109
13	CDK6	H	200a	110
14	CDKN1B (p27/kip1)	H	200b	111
15	CFL2	M	200c	112
		H	200b	109
16	c-Maf	M	200c	113
17	c-MYB	H	200b, 200c, 429	114
18	C-MYC	H	429	115
19	CREB1	H	200b	116
20	CTNNB1 (β-CATENIN)	R	200a	117-119
		H	200a	120
		H	200a	121

Table 3 (cont.). Published miR-200 target genes. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown.

21	CUL3	H	141	105
22	CXCL1	H	200a, 200b	122
23	CXCL12β	H, M	141	123
24	CYCLIN E2	H	200a, 200b, 200c	124
25	DLC-1	H	141	125
26	Dlx5	M	141, 200a	126
27	DNMT3A	H	200b, 200c	127
28	DNMT3B	H	200b, 200c	127
29	E2f3	M	200c	128
30	eIF4E	H	141	129
31	EPHA2	H	200a	130
32	EPHRIN-A1 (EFNA1)	H	200c	131
33	ERRFI-1	H	200c	132
34	ETAR (EDNRA)	H	200c	133
35	ETS-1	M H	200c 200b	134 135
36	FAP-1 (PTPN13)	H	200c	136
37	FBLN5	H	200c	137
38	FHOD1	H	200c	132
39	FLK1 (KDR, VEGFR2)	M M H H	200c 200b 200b 200c	134 138 139 140
40	FLT1 (VEGFR1)	H H H	200b 200c 200c	139 141 142
41	FN1	H M H	200c 200c/141 200b	92 143 144
42	FOG2 (ZFPM2)	H	200a, 200b, 200c 141, 429	145
43	FTH1 (Ferritin)	H	200b	146
44	GATA2	H	200b	138
45	GATA4	H H	200b 200c	147 148
46	GNA13	H	200a 141	149
47	Grb2	M	200a	111,150
48	HDAC4	H	200a	151
49	HDGF	H	141	152
50	IKKB (Ikbkb)	H	200c	141

Table 3 (cont.). Published miR-200 target genes. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown.

51	IL-8 (Cxcl15)	H	200a, 200b	122
52	JAGGED1 (Jag1)	H	200c, 141	153-155
53	KEAP1	H H M	141 200a 200c/141	156 157 143
54	KINDLIN-2 (FERMT2)	H	200b	109
55	KLF4	H	200b	96
56	KLF9	H	200c	141
57	KRAS	H	200c	158
58	LEPR	H	200c	92
59	LPAR1	H	200c	133
60	Lrp1	M	200c	112
61	MAML2	H	200c	154
62	MAML3	H	200c	154
63	MAP4K4	H	141	159
64	MAPK14 (P38A)	H	200a, 141	160
65	MARCKS	H	200c	161
66	MKK4	H	141	162
67	MSN	H H	200c 200b	92 163
68	MUC16	H	200c	164
69	MUC4	H	200c	164
70	NCAM1	H	200c	165
71	Noggin (Nog)	M	200c	166
72	NOTCH1	H H	200b, 200c 200b	167 168
73	NOXA	H	200c	169
74	Nrp1	M	200a	108
75	NTF3	H	200c	170
76	NTRK2 (TRKB)	H	200c	92
77	ONECUT2	H	429	171
78	OREBP	H	200b	172
79	OXR1	H	200b	173
80	P53	H	200a	174
81	PAF	H	200b	109
82	PIN1	H	200b	175
83	Pkd1 (polycystic kidney disease 1)	M	200a, 200b	176

Table 3 (cont.). Published miR-200 target genes. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown.

84	PLCG1	H	200b,200c,429	177
85	PPARA	H	141	178
86	PPM1F	H	200c	179
87	PRKAR2B	H	200b	180
88	PSMD1	H	200b	181
89	PSMD1	H	200b	181
90	PTEN	H	141	102
		H	200a	182
		H	200c	183
		M	200a	184
91	RAB18	H	429	185
		H	200b	186-187
92	RAB21	H	200b	186-187
93	RAB23	H	200b	186-187
94	RAB3B	H	200b	186-187
95	RBBP4	H	429	188
96	RHOA	H	200c	133
97	RND3 (RhoE)	H	200b	189
		H	200b	111
		H	200c	190
98	ROCK2	H	200b, 200c	191
99	SEC23A	H	200c	107,192
		M	200c	112
100	Serca2b (Atp2a2)	M	200b	193
101	SHP (NR0B2)	H	141	194
102	SIRT1	H	200a	195
103	Slc25a3	M	141	196
104	Slug (Snai2)	M	200b	197
105	SMAD2	H	141, 200c	198
		M	200b	199
		H	200b	200
106	Snail (Snai1)	M	200b	199
107	SOX17	H	141, 200a	155
		H	200a	201
108	SOX2	H	429	202
		M	200c	128
109	SP1	H	200b, 200c	127
		H	429	94
110	Srf	R	200b	203
111	STAT5B	H	200a	204
112	SUZ12	H	200b	96,205
		H	200b, 200c	191
113	TBK1	H	200c	206

Table 3 (cont.). Published miR-200 target genes. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown.

114	TGF-B2	H	141	[207]
		R	200a	[208]
		M	200c/141	[143]
		R	200a	[118]
115	TGFBR1	H	141, 200a	[198]
116	THBS1	H	200a	[209]
117	TIAM1	H	141	[210]
118	TIMP2	H	200b	[211]
119	TIMP2	H	200c	[137]
120	TM4SF1	H	141	[212]
121	TUBB3	H	200c	[213]
		H	200c	[214]
122	UBAP1	H	141	[102]
123	UBASH3B	H	200a	[215]
124	UBQLN1	H	200b	[181]
125	UBQLN1	H	200b	[181]
126	VEGF-A	H, M	200b	[216]
		H, R	200b	[217]
		H	200b	[139]
		H	200c	[141]
		H	200c	[137]
127	WAVE3 (WASF3)	H	200b	[218]
128	Wnt1	M	200b	[219]
129	XIAP	H	200b, 200c, 429	[93]
130	YAP1	H	200a	[220]
		H	141	[221-222]
131	YWHAG (14-3-3γ)	H	141	[223]
132	ZEB1	H	200a, 200b	[224]
		M	200b, 200c	[225]
133	ZEB2 (SIP1)	H	200a, 200b	[224]
		M	200b, 200c	[225]
134	ZMPSTE24	H	141	[226]
135	ZNF217	H	200c	[227]

3.2. miR-200 family role in EMT and metastasis

Because of its role as a master EMT regulator, miR-200 has been extensively studied. Besides targeting EMT-inducers Zeb1 and Zeb2, miR-200 also targets other important EMT activators such as Ets1, Snail1 (Snail) and Snail2 (Slug) (Table 3). miR-200 has also been shown to regulate cell-cell adhesion[207,224,228], cancer stem cell self-renewal and differentiation[96,229-230], cell division and apoptosis[136,147,177,189] and chemoresistance[93,231] (Figure 10).

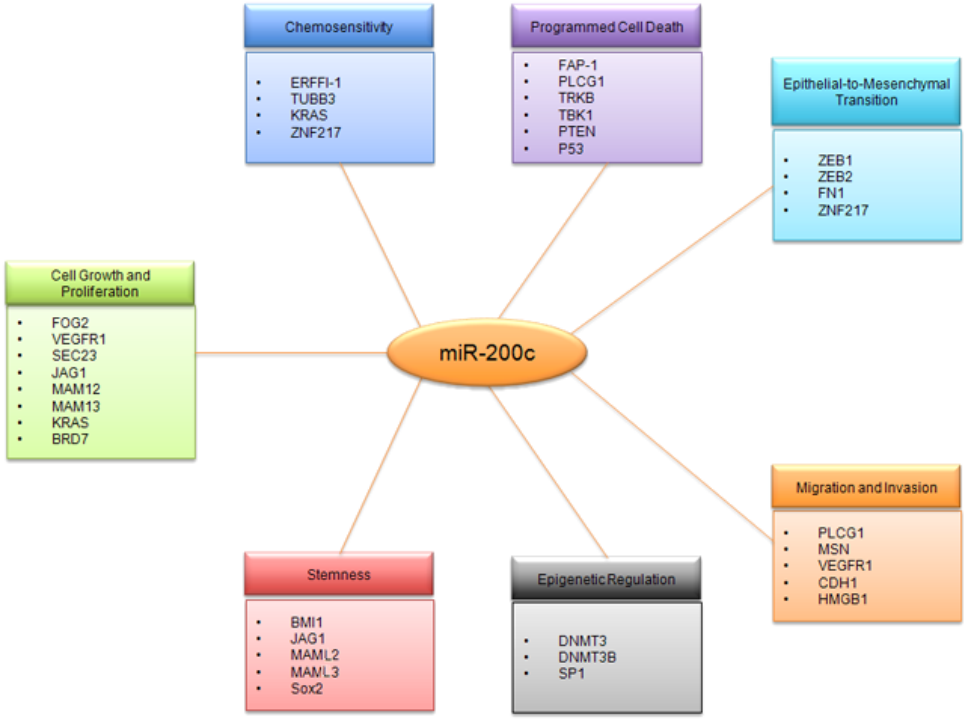


Figure 10. miR-200c targets several genes regulating numerous processes involved in cancer development and progression. Figure and legend reproduced from Mutlu M SÖ, Raza U, Eyüpoglu E, Yurdusev E, Sahin Ö. Atlas Genet Cytogenet Oncol Haematol 2015[88].

miR-200c accounts for 93% of the mature miR-200b/c/429 subfamily expression in human breast tumors[232]. Due to the dominant presence of miR-200c over its other subfamily members and given the expected overlap

between miR-200b/c/429 and miR-200a/141 subfamily target genes, we chose for this project to focus on miR-200c as a representative member of the miR-200 family.

Even though miR-200 was shown to regulate metastasis formation, its effect is controversial as miR-200 expression enhances metastasis in some cancer models,[112,202,225,233-234] but reduces it in others[95,186,235]. Different cancer cell models may have different factors that alter miR-200's role in metastasis. Additionally, the microenvironment surrounding the primary tumor may also contribute to these differences. Extracellular cytokines at the invasive front, such as TGF-β, may activate miR-200 repressors such as Zeb1/2 that repress miR-200 expression at the transcriptional level. Finally, the microenvironment at secondary metastatic sites may also cooperate or counteract miR-200's colonization-promoting effect.

Circulating miR-200 levels can also be used as a prognostic tool in oncology, as high levels of miR-200 in the bloodstream of cancer patients are associated with poor prognosis in several cancers, including ovarian, prostate, pancreatic and metastatic colorectal cancers[236-239].

4. Thesis Scope

The main goals of the work within this thesis were to provide the first genome-wide map of miR-200c target genes and investigate their biological role. Although miR-200 family was already known to regulate EMT, a genome-wide list of its targets and a comprehensive map of miR-200-regulated pathways was yet to be obtained.

To achieve these main goals an integrated framework was developed that consisted on coupling a high-throughput method for miRNA target discovery with bioinformatic analyses. This method identified hundreds of

miR-200c putative targets. This genome-wide list of miR-200c target genes was then bioinformatically analyzed to provide a global list of miR-200-regulated pathways. Some of these putative target genes were selected for experimental validation. Results show that this high-throughput method for miRNA target discovery has a low false-positive rate. Additionally, biochemical assays were performed to study the role of some novel target genes in miR-200-dependent processes. Some of these target genes had novel and unexpected biological functions.

Overall, the framework developed in this thesis allowed us to generate the most complete genome-wide map of miR-200-regulated pathways to date and discover new genes involved in EMT. Moreover, this framework can be used to identify the genome-wide targets of other miRNAs.

Importantly, this thesis makes a significant contribution to the study of both miR-200 and EMT and will be the springboard for future mechanistic and biological investigations of the molecular basis for their actions.

5. References

1 Thomas M, Lieberman J, Lal A. Desperately seeking microRNA targets. *Nat Struct Mol Biol* 2010; **17:** 1169-1174.

2 Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* 2009; **11:** 228-234.

3 Ardekani AM, Naeini MM. The Role of MicroRNAs in Human Diseases. *Avicenna J Med Biotechnol* 2010; **2:** 161-179.

4 Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 2009; **19:** 92-105.

5 Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH *et al*. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* 2004; **23:** 4051-4060.

6 Hata A, Lieberman J. Dysregulation of microRNA biogenesis and gene silencing in cancer. *Sci Signal* 2015; **8:** re3.

7 Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 2005; **11:** 241-247.

8 Kim VN, Nam JW. Genomics of microRNA. *Trends Genet* 2006; **22:** 165-173.

9 Yang JS, Lai EC. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol Cell* 2011; **43:** 892-903.

10 Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 2002; **297:** 2056-2060.

11 Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ *et al*. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 2004; **305:** 1437-1441.

12 Jones-Rhoades MW, Bartel DP, Bartel B. MicroRNAS and their regulatory roles in plants. *Annu Rev Plant Biol* 2006; **57:** 19-53.

13 Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; **136:** 215-233.

14 Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 2008; **9:** 22-32.

15 Zekri L, Huntzinger E, Heimstadt S, Izaurralde E. The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol Cell Biol* 2009; **29:** 6220-6231.

16 Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat Rev Genet* 2015; **16:** 421-433.

17 Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell* 2003; **115:** 787-798.

18 Nielsen CB, Shomron N, Sandberg R, Hornstein E, Kitzman J, Burge CB. Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 2007; **13:** 1894-1910.

19 Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA* 2004; **10:** 1507-1517.

20 Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; **27:** 91-105.

21 Ahmadi H, Ahmadi A, Azimzadeh-Jamalkandi S, Shoorehdeli MA, Salehzadeh-Yazdi A, Bidkhori G *et al*. HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens. *Genomics* 2013; **101:** 94-100.

22 Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE *et al*. Comprehensive discovery of endogenous Argonaute binding sites in Caenorhabditis elegans. *Nat Struct Mol Biol* 2010; **17:** 173-179.

23 Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009; **460:** 479-486.

24 Lal A, Navarro F, Maher CA, Maliszewski LE, Yan N, O'Day E *et al*. miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. *Mol Cell* 2009; **35:** 610-625.

25 Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 2010; **38:** 789-802.

26 Johnson CD, Esquela-Kerscher A, Stefani G, Byrom M, Kelnar K, Ovcharenko D *et al*. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res* 2007; **67:** 7713-7722.

27 Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. *Genes Dev* 2004; **18:** 132-137.

28 Tarang S, Weston MD. Macros in microRNA target identification: a comparative analysis of in silico, in vitro, and in vivo approaches to microRNA target identification. *RNA Biol* 2014; **11:** 324-333.

29 Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 2008; **320:** 1643-1647.

30 Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* 2009; **106:** 7028-7033.

31 Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 2009; **138:** 673-684.

32 Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C *et al*. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008; **456:** 470-476.

33 Nicolas FE. Experimental validation of microRNA targets using a luciferase reporter system. *Methods Mol Biol* 2011; **732:** 139-152.

34 Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. *Nature* 2008; **455:** 64-71.

35 Hu R, Wallace J, Dahlem TJ, Grunwald DJ, O'Connell RM. Targeting human microRNA genes using engineered Tal-effector nucleases (TALENs). *PLoS One* 2013; **8:** e63074.

36 Krutzfeldt J, Rajewsky N, Braich R, Rajeev KG, Tuschl T, Manoharan M *et al*. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* 2005; **438:** 685-689.

37 Kloosterman WP, Wienholds E, de Bruijn E, Kauppinen S, Plasterk RH. In situ detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat Methods* 2006; **3:** 27-29.

38 Choi WY, Giraldez AJ, Schier AF. Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430. *Science* 2007; **318:** 271-274.

39 Ebert MS, Neilson JR, Sharp PA. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat Methods* 2007; **4:** 721-726.

40 Khan AA, Betel D, Miller ML, Sander C, Leslie CS, Marks DS. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotechnol* 2009; **27:** 549-555.

41 Stenvang J, Petri A, Lindow M, Obad S, Kauppinen S. Inhibition of microRNA function by antimiR oligonucleotides. *Silence* 2012; **3:** 1.

42 Beitzinger M, Peters L, Zhu JY, Kremmer E, Meister G. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol* 2007; **4:** 76-84.

43 Easow G, Teleman AA, Cohen SM. Isolation of microRNA targets by miRNP immunopurification. *RNA* 2007; **13:** 1198-1204.

44 Landthaler M, Gaidatzis D, Rothballer A, Chen PY, Soll SJ, Dinic L *et al*. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* 2008; **14:** 2580-2596.

45 Hong X, Hammell M, Ambros V, Cohen SM. Immunopurification of Ago1 miRNPs selects for a distinct class of microRNA targets. *Proc Natl Acad Sci U S A* 2009; **106:** 15085-15090.

46 Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G *et al*. A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci U S A* 2007; **104:** 19291-19296.

47 Hendrickson DG, Hogan DJ, Herschlag D, Ferrell JE, Brown PO. Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS One* 2008; **3:** e2126.

48 Su H, Trombly MI, Chen J, Wang X. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev* 2009; **23:** 304-317.

49 Ipsaro JJ, Joshua-Tor L. From guide to target: molecular insights into eukaryotic RNA-interference machinery. *Nat Struct Mol Biol* 2015; **22:** 20-28.

50 Maniataki E, Mourelatos Z. Human mitochondrial tRNAMet is exported to the cytoplasm and associates with the Argonaute 2 protein. *RNA* 2005; **11:** 849-852.

51 Zhang X, Graves PR, Zeng Y. Stable Argonaute2 overexpression differentially regulates microRNA production. *Biochim Biophys Acta* 2009; **1789:** 153-159.

52 Diederichs S, Haber DA. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell* 2007; **131:** 1097-1108.

53 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P *et al*. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010; **141:** 129-141.

54 Sugimoto Y, Konig J, Hussain S, Zupan B, Curk T, Frye M *et al*. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 2012; **13:** R67.

55 Lal A, Thomas MP, Altschuler G, Navarro F, O'Day E, Li XL *et al*. Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet* 2011; **7:** e1002363.

56 Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 2008; **30:** 460-471.

57 Krishnan K, Steptoe AL, Martin HC, Wani S, Nones K, Waddell N *et al*. MicroRNA-182-5p targets a network of genes involved in DNA repair. *RNA* 2013; **19:** 230-242.

58 Kang H, Davis-Dusenbery BN, Nguyen PH, Lal A, Lieberman J, Van Aelst L *et al*. Bone morphogenetic protein 4 promotes vascular smooth muscle contractility by activating microRNA-21

(miR-21), which down-regulates expression of family of dedicator of cytokinesis (DOCK) proteins. *J Biol Chem* 2012; **287:** 3976-3986.

59 Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S *et al*. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol* 2011; **12:** R126.

60 Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W *et al*. Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. *Cell Rep* 2014; **8:** 1225-1239.

61 Kozomara A, Hunt S, Ninova M, Griffiths-Jones S, Ronshaugen M. Target repression induced by endogenous microRNAs: large differences, small effects. *PLoS One* 2014; **9:** e104286.

62 Subramanian M, Li XL, Hara T, Lal A. A biochemical approach to identify direct microRNA targets. *Methods Mol Biol* 2015; **1206:** 29-37.

63 Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest* 2009; **119:** 1420-1428.

64 Duband JL, Monier F, Delannet M, Newgreen D. Epithelium-mesenchyme transition during neural crest development. *Acta Anat (Basel)* 1995; **154:** 63-78.

65 Kang P, Svoboda KK. Epithelial-mesenchymal transformation during craniofacial development. *J Dent Res* 2005; **84:** 678-690.

66 Yang J, Weinberg RA. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* 2008; **14:** 818-829.

67 Thiery JP. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2002; **2:** 442-454.

68 Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell* 2009; **139:** 871-890.

69 Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. *J Clin Invest* 2009; **119:** 1429-1437.

70 Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. *Nat Rev Cancer* 2005; **5:** 591-602.

71 Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *J Cell Biol* 2006; **172:** 973-981.

72 Wells A, Yates C, Shepard CR. E-cadherin as an indicator of mesenchymal to epithelial reverting transitions during the metastatic seeding of disseminated carcinomas. *Clin Exp Metastasis* 2008; **25:** 621-628.

73 Wan L, Pantel K, Kang Y. Tumor metastasis: moving new biological insights into the clinic. *Nat Med* 2013; **19:** 1450-1464.

74 Wu Y, Zhang X, Salmon M, Lin X, Zehner ZE. TGFbeta1 regulation of vimentin gene expression during differentiation of the C2C12 skeletal myogenic cell line requires Smads, AP-1 and Sp1 family members. *Biochim Biophys Acta* 2007; **1773:** 427-439.

75 Thuault S, Tan EJ, Peinado H, Cano A, Heldin CH, Moustakas A. HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. *J Biol Chem* 2008; **283:** 33437-33446.

76 Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA. Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res* 2008; **68:** 3645-3654.

77 Wrighton KH, Lin X, Feng XH. Phospho-control of TGF-beta superfamily signaling. *Cell Res* 2009; **19:** 8-20.

78 Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* 2003; **425:** 577-584.

79 Massague J, Wotton D. Transcriptional control by the TGF-beta/Smad signaling system. *EMBO J* 2000; **19:** 1745-1754.

80 Vincent T, Neve EP, Johnson JR, Kukalev A, Rojo F, Albanell J *et al*. A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition. *Nat Cell Biol* 2009; **11:** 943-950.

81 Postigo AA. Opposing functions of ZEB proteins in the regulation of the TGFbeta/BMP signaling pathway. *EMBO J* 2003; **22:** 2443-2452.

82 Verschueren K, Remacle JE, Collart C, Kraft H, Baker BS, Tylzanowski P *et al*. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem* 1999; **274:** 20489-20498.

83 Luo K. Ski and SnoN: negative regulators of TGF-beta signaling. *Curr Opin Genet Dev* 2004; **14:** 65-70.

84 Pessah M, Prunier C, Marais J, Ferrand N, Mazars A, Lallemand F *et al*. c-Jun interacts with the corepressor TG-interacting factor (TGIF) to suppress Smad2 transcriptional activity. *Proc Natl Acad Sci U S A* 2001; **98:** 6198-6203.

85 Lan HY. Smad7 as a therapeutic agent for chronic kidney diseases. *Front Biosci* 2008; **13:** 4984-4992.

86 Li JH, Zhu HJ, Huang XR, Lai KN, Johnson RJ, Lan HY. Smad7 inhibits fibrotic effect of TGF-Beta on renal tubular epithelial cells by blocking Smad2 activation. *J Am Soc Nephrol* 2002; **13:** 1464-1472.

87 Arora K, Warrior R. A new Smurf in the village. *Dev Cell* 2001; **1:** 441-442.

88 Mutlu M SÖ, Raza U, Eyüpoglu E, Yurdusev E, Sahin Ö. MIR200C (microRNA 200c). *Atlas Genet Cytogenet Oncol Haematol* 2015; **19:** 270-285.

89 Perdigao-Henriques R, Petrocca F, Altschuler G, Thomas MP, Le MT, Tan SM *et al*. miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes. *Oncogene* 2015.

90 Senanayake U, Das S, Vesely P, Alzoughbi W, Frohlich LF, Chowdhury P *et al*. miR-192, miR-194, miR-215, miR-200c and miR-141 are downregulated and their common target ACVR2B is strongly expressed in renal childhood neoplasms. *Carcinogenesis* 2012; **33:** 1014-1021.

91 Wu Y, Xiao Y, Ding X, Zhuo Y, Ren P, Zhou C *et al*. A miR-200b/200c/429-binding site polymorphism in the 3' untranslated region of the AP-2alpha gene is associated with cisplatin resistance. *PLoS One* 2011; **6:** e29043.

92 Howe EN, Cochrane DR, Richer JK. Targets of miR-200c mediate suppression of cell motility and anoikis resistance. *Breast Cancer Res* 2011; **13:** R45.

93 Zhu W, Xu H, Zhu D, Zhi H, Wang T, Wang J *et al*. miR-200bc/429 cluster modulates multidrug resistance of human cancer cell lines by targeting BCL2 and XIAP. *Cancer Chemother Pharmacol* 2012; **69:** 723-731.

94 Wang Y, Li M, Zang W, Ma Y, Wang N, Li P *et al*. MiR-429 up-regulation induces apoptosis and suppresses invasion by targeting Bcl-2 and SP-1 in esophageal carcinoma. *Cell Oncol (Dordr)* 2013; **36:** 385-394.

95 Sun L, Yao Y, Liu B, Lin Z, Lin L, Yang M *et al*. MiR-200b and miR-15b regulate chemotherapy-induced epithelial-mesenchymal transition in human tongue cancer cells by targeting BMI1. *Oncogene* 2012; **31:** 432-445.

96 Iliopoulos D, Lindahl-Allen M, Polytarchou C, Hirsch HA, Tsichlis PN, Struhl K. Loss of miR-200 Inhibition of Suz12 Leads to Polycomb-Mediated Repression Required for the Formation and Maintenance of Cancer Stem Cells. *Mol Cell* 2010; **39:** 761-772.

97 Yin J, Zheng G, Jia X, Zhang Z, Zhang W, Song Y *et al*. A Bmi1-miRNAs cross-talk modulates chemotherapy response to 5-fluorouracil in breast cancer cells. *PLoS One* 2013; **8:** e73268.

98 Lo WL, Yu CC, Chiou GY, Chen YW, Huang PI, Chien CS *et al*. MicroRNA-200c attenuates tumour growth and metastasis of presumptive head and neck squamous cell carcinoma stem cells. *J Pathol* 2011; **223:** 482-495.

99 Wellner U, Schubert J, Burk UC, Schmalhofer O, Zhu F, Sonntag A *et al*. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol* 2009; **11:** 1487-1495.

100 Dimri M, Carroll JD, Cho JH, Dimri GP. microRNA-141 regulates BMI1 expression and induces senescence in human diploid fibroblasts. *Cell Cycle* 2013; **12:** 3537-3546.

101 Cui J, Cheng Y, Zhang P, Sun M, Gao F, Liu C *et al*. Down Regulation of miR200c Promotes Radiation-Induced Thymic Lymphoma by Targeting BMI1. *J Cell Biochem* 2013; **115:** 1033-1042.

102 Zhang L, Deng T, Li X, Liu H, Zhou H, Ma J *et al*. microRNA-141 is involved in a nasopharyngeal carcinoma-related genes network. *Carcinogenesis* 2010; **31:** 559-566.

103 Park YA, Lee JW, Choi JJ, Jeon HK, Cho Y, Choi C *et al*. The interactions between MicroRNA-200c and BRD7 in endometrial carcinoma. *Gynecol Oncol* 2012; **124:** 125-133.

104 Kim Y, Park D, Kim H, Choi M, Lee H, Lee YS *et al*. miR-200b and cancer/testis antigen CAGE form a feedback loop to regulate the invasion and tumorigenic and angiogenic responses of a cancer cell line to microtubule-targeting drugs. *J Biol Chem* 2013; **288:** 36502-36518.

105 Tang W, Qin J, Tang J, Zhang H, Zhou Z, Li B *et al*. Aberrant Reduction of MiR-141 Increased CD47/CUL3 in Hirschsprung's Disease. *Cell Physiol Biochem* 2013; **32:** 1655-1667.

106 Yu XY, Zhang Z, Liu J, Zhan B, Kong CZ. MicroRNA-141 is downregulated in human renal cell carcinoma and regulates cell survival by targeting CDC25B. *Onco Targets Ther* 2013; **6:** 349-354.

107 Luo D, Wilson JM, Harvel N, Liu J, Pei L, Huang S *et al*. A systematic evaluation of miRNA:mRNA interactions involved in the migration and invasion of breast cancer cells. *J Transl Med* 2013; **11:** 57.

108 Lin CH, Jackson AL, Guo J, Linsley PS, Eisenman RN. Myc-regulated microRNAs attenuate embryonic stem cell differentiation. *EMBO J* 2009; **28:** 3157-3170.

109 Zhang HF, Zhang K, Liao LD, Li LY, Du ZP, Wu BL *et al*. miR-200b suppresses invasiveness and modulates the cytoskeletal and adhesive machinery in esophageal squamous cell carcinoma cells via targeting Kindlin-2. *Carcinogenesis* 2014; **35:** 292-301.

110 Xiao F, Zhang W, Zhou L, Xie H, Xing C, Ding S *et al*. microRNA-200a is an independent prognostic factor of hepatocellular carcinoma and induces cell cycle arrest by targeting CDK6. *Oncol Rep* 2010; **30:** 2203-2210.

111 Fu Y, Liu X, Zhou N, Du L, Sun Y, Zhang X *et al*. MicroRNA-200b stimulates tumour growth in TGFBR2-null colorectal cancers by negatively regulating p27/kip1. *J Cell Physiol* 2013; **229:** 772-782.

112 Korpal M, Ell BJ, Buffa FM, Ibrahim T, Blanco MA, Celia-Terrassa T *et al*. Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* 2011; **17:** 1101-1108.

113 Klein D, Misawa R, Bravo-Egana V, Vargas N, Rosero S, Piroso J *et al*. MicroRNA Expression in Alpha and Beta Cells of Human Pancreatic Islets. *PLoS One* 2013; **8:** e55064.

114 Cesi V, Casciati A, Sesti F, Tanno B, Calabretta B, Raschella G. TGFbeta-induced c-Myb affects the expression of EMT-associated genes and promotes invasion of ER+ breast cancer cells. *Cell Cycle* 2011; **10:** 4149-4161.

115 Sun T, Wang C, Xing J, Wu D. miR-429 modulates the expression of c-myc in human gastric carcinoma cells. *Eur J Cancer* 2011; **47:** 2552-2559.

116 Peng B, Hu S, Jun Q, Luo D, Zhang X, Zhao H *et al*. MicroRNA-200b targets CREB1 and suppresses cell growth in human malignant glioma. *Mol Cell Biochem* 2013; **379:** 51-58.

117 Wu X, Zhao B, Li W, Chen Y, Liang R, Li L *et al*. MiR-200a is involved in rat epididymal development by targeting beta-catenin mRNA. *Acta Biochim Biophys Sin (Shanghai)* 2012; **44:** 233-240.

118 Sun X, He Y, Ma TT, Huang C, Zhang L, Li J. Participation of miR-200a in TGF-beta1-mediated hepatic stellate cell activation. *Mol Cell Biochem* 2014; **388:** 11-23.

119 Liu J, Ruan B, You N, Huang Q, Liu W, Dang Z *et al*. Downregulation of miR-200a induces EMT phenotypes and CSC-like signatures through targeting the beta-catenin pathway in hepatic oval cells. *PLoS One* 2013; **8:** e79409.

120 Saydam O, Shen Y, Wurdinger T, Senol O, Boke E, James MF *et al*. Downregulated microRNA-200a in meningiomas promotes tumor growth by reducing E-cadherin and activating the Wnt/beta-catenin signaling pathway. *Mol Cell Biol* 2009; **29:** 5923-5940.

121 Su J, Zhang A, Shi Z, Ma F, Pu P, Wang T *et al*. MicroRNA-200a suppresses the Wnt/beta-catenin signaling pathway by interacting with beta-catenin. *Int J Oncol* 2012; **40:** 1162-1170.

122 Pecot CV, Rupaimoole R, Yang D, Akbani R, Ivan C, Lu C *et al*. Tumour angiogenesis regulation by the miR-200 family. *Nat Commun* 2013; **4:** 2427.

123 Huang Z, Shi T, Zhou Q, Shi S, Zhao R, Shi H *et al*. miR-141 Regulates colonic leukocytic trafficking by targeting CXCL12beta during murine colitis and human Crohn's disease. *Gut* 2013.

124 Cai J, Liu X, Cheng J, Li Y, Huang X, Li Y *et al*. MicroRNA-200 is commonly repressed in conjunctival MALT lymphoma, and targets cyclin E2. *Graefes Arch Clin Exp Ophthalmol* 2012; **250:** 523-531.

125 Banaudha K, Kaliszewski M, Korolnek T, Florea L, Yeung ML, Jeang KT *et al*. MicroRNA silencing of tumor suppressor DLC-1 promotes efficient hepatitis C virus replication in primary human hepatocytes. *Hepatology* 2011; **53:** 53-61.

126 Itoh T, Nozawa Y, Akao Y. MicroRNA-141 and -200a are involved in bone morphogenetic protein-2-induced mouse pre-osteoblast differentiation by targeting distal-less homeobox 5. *J Biol Chem* 2009; **284:** 19272-19279.

127 Tang H, Deng M, Tang Y, Xie X, Guo J, Kong Y *et al*. miR-200b and miR-200c as Prognostic Factors and Mediators of Gastric Cancer Cell Progression. *Clin Cancer Res* 2013; **19:** 5602-5612.

128 Peng C, Li N, Ng YK, Zhang J, Meier F, Theis FJ *et al*. A unilateral negative feedback loop between miR-200 microRNAs and Sox2/E2F3 controls neural progenitor cell-cycle exit and differentiation. *J Neurosci* 2012; **32:** 13292-13308.

129 Ho BC, Yu SL, Chen JJ, Chang SY, Yan BS, Hong QS *et al*. Enterovirus-induced miR-141 contributes to shutoff of host protein translation by targeting the translation initiation factor eIF4E. *Cell Host Microbe* 2011; **9:** 58-69.

130 Sun Q, Zou X, Zhang T, Shen J, Yin Y, Xiang J. The role of miR-200a in vasculogenic mimicry and its clinical significance in ovarian cancer. *Gynecol Oncol* 2014; **132:** 730-738.

131 Li Y, Nie Y, Cao J, Tu S, Lin Y, Du Y *et al*. G-A variant in miR-200c binding site of EFNA1 alters susceptibility to gastric cancer. *Mol Carcinog* 2012; **53:** 219-229.

132 Adam L, Zhong M, Choi W, Qi W, Nicoloso M, Arora A *et al*. miR-200 expression regulates epithelial-to-mesenchymal transition in bladder cancer cells and reverses resistance to epidermal growth factor receptor therapy. *Clin Cancer Res* 2009; **15:** 5060-5072.

133 Luna C, Li G, Huang J, Qiu J, Wu J, Yuan F *et al*. Regulation of trabecular meshwork cell contraction and intraocular pressure by miR-200c. *PLoS One* 2012; **7:** e51688.

134 Gill JG, Langer EM, Lindsley RC, Cai M, Murphy TL, Murphy KM. Snail promotes the cell-autonomous generation of Flk1(+) endothelial cells through the repression of the microRNA-200 family. *Stem Cells Dev* 2012; **21:** 167-176.

135 Chan YC, Khanna S, Roy S, Sen CK. miR-200b targets Ets-1 and is down-regulated by hypoxia to induce angiogenic response of endothelial cells. *J Biol Chem* 2011; **286:** 2047-2056.

136 Schickel R, Park SM, Murmann AE, Peter ME. mir-200c Regulates Induction of Apoptosis through CD95 by Targeting FAP-1. *Mol Cell* 2010; **38:** 908-915.

137 Chuang TD, Panda H, Luo X, Chegini N. miR-200c is aberrantly expressed in leiomyomas in an ethnic-dependent manner and targets ZEBs, VEGFA, TIMP2, and FBLN5. *Endocr Relat Cancer* 2012; **19:** 541-556.

138 Chan YC, Roy S, Khanna S, Sen CK. Downregulation of endothelial microRNA-200b supports cutaneous wound angiogenesis by desilencing GATA binding protein 2 and vascular endothelial growth factor receptor 2. *Arterioscler Thromb Vasc Biol* 2012; **32:** 1372-1382.

139 Choi YC, Yoon S, Jeong Y, Yoon J, Baek K. Regulation of vascular endothelial growth factor signaling by miR-200b. *Mol Cells* 2011; **32:** 77-82.

140 Shi L, Zhang S, Wu H, Zhang L, Dai X, Hu J *et al*. MiR-200c increases the radiosensitivity of non-small-cell lung cancer cell line A549 by targeting VEGF-VEGFR2 pathway. *PLoS One* 2013; **8:** e78344.

141 Panda H, Pelakh L, Chuang TD, Luo X, Bukulmez O, Chegini N. Endometrial miR-200c is altered during transformation into cancerous states and targets the expression of ZEBs, VEGFA, FLT1, IKKbeta, KLF9, and FBLN5. *Reprod Sci* 2012; **19:** 786-796.

142 Mezquita B, Mezquita J, Barrot C, Carvajal S, Pau M, Mezquita P *et al*. A truncated-Flt1 isoform of breast cancer cells is upregulated by Notch and downregulated by retinoic acid. *J Cell Biochem* 2014; **115:** 52-61.

143 Wei J, Zhang Y, Luo Y, Wang Z, Bi S, Song D *et al*. Aldose reductase regulates miR-200a-3p/141-3p to coordinate Keap1-Nrf2, Tgfbeta1/2, and Zeb1/2 signaling in renal mesangial cells and the renal cortex of diabetic mice. *Free Radic Biol Med* 2014; **67:** 91-102.

144 Tang O, Chen XM, Shen S, Hahn M, Pollock CA. MiRNA-200b represses transforming growth factor-beta1-induced EMT and fibronectin expression in kidney proximal tubular cells. *Am J Physiol Renal Physiol* 2013; **304:** F1266-1273.

145 Hyun S, Lee JH, Jin H, Nam J, Namkoong B, Lee G *et al*. Conserved MicroRNA miR-8/miR-200 and its target USH/FOG2 control growth by regulating PI3K. *Cell* 2009; **139:** 1096-1108.

146 Shpyleva SI, Tryndyak VP, Kovalchuk O, Starlard-Davenport A, Chekhun VF, Beland FA *et al*. Role of ferritin alterations in human breast cancer cells. *Breast Cancer Res Treat* 2011; **126:** 63-71.

147 Yao CX, Wei QX, Zhang YY, Wang WP, Xue LX, Yang F *et al*. miR-200b targets GATA-4 during cell growth and differentiation. *RNA Biol* 2013; **10:** 465-480.

148 Huang HN, Chen SY, Hwang SM, Yu CC, Su MW, Mai W *et al*. miR-200c and GATA binding protein 4 regulate human embryonic stem cell renewal and differentiation. *Stem Cell Res* 2013; **12:** 338-353.

149 Rasheed SA, Teo CR, Beillard EJ, Voorhoeve PM, Casey PJ. MicroRNA-182 and microRNA-200a control G-protein subunit alpha-13 (GNA13) expression and cell invasion synergistically in prostate cancer cells. *J Biol Chem* 2013; **288:** 7986-7995.

150 Liu Y, Liu Q, Jia W, Chen J, Wang J, Ye D *et al*. MicroRNA-200a regulates Grb2 and suppresses differentiation of mouse embryonic stem cells into endoderm and mesoderm. *PLoS One* 2013; **8:** e68990.

151 Yuan JH, Yang F, Chen BF, Lu Z, Huo XS, Zhou WP *et al*. The histone deacetylase 4/SP1/microrna-200a regulatory network contributes to aberrant histone acetylation in hepatocellular carcinoma. *Hepatology* 2011; **54:** 2025-2035.

152 Chen B, Huang T, Jiang J, Lv L, Li H, Xia S. miR-141 suppresses proliferation and motility of gastric cancer cells by targeting HDGF. *Mol Cell Biochem* 2014; **388:** 211-218.

153 Vallejo DM, Caparros E, Dominguez M. Targeting Notch signalling by the conserved miR-8/200 microRNA family in development and cancer cells. *EMBO J* 2011; **30:** 756-769.

154 Brabletz S, Bajdak K, Meidhof S, Burk U, Niedermann G, Firat E *et al*. The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. *EMBO J* 2011; **30:** 770-782.

155 Jia Y, Yang Y, Zhan Q, Brock MV, Zheng X, Yu Y *et al*. Inhibition of SOX17 by microRNA 141 and methylation activates the WNT signaling pathway in esophageal cancer. *J Mol Diagn* 2012; **14:** 577-585.

156 van Jaarsveld MT, Helleman J, Boersma AW, van Kuijk PF, van Ijcken WF, Despierre E *et al*. miR-141 regulates KEAP1 and modulates cisplatin sensitivity in ovarian cancer cells. *Oncogene* 2013; **32:** 4284-4293.

157 Eades G, Yang M, Yao Y, Zhang Y, Zhou Q. miR-200a regulates Nrf2 activation by targeting Keap1 mRNA in breast cancer cells. *J Biol Chem* 2011; **286:** 40725-40733.

158 Kopp F, Wagner E, Roidl A. The proto-oncogene KRAS is targeted by miR-200c. *Oncotarget* 2014; **5:** 185-195.

159 Zhao G, Wang B, Liu Y, Zhang JG, Deng SC, Qin Q *et al*. miRNA-141, downregulated in pancreatic cancer, inhibits cell proliferation and invasion by directly targeting MAP4K4. *Mol Cancer Ther* 2013; **12:** 2569-2580.

160 Mateescu B, Batista L, Cardon M, Gruosso T, de Feraudy Y, Mariani O *et al*. miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat Med* 2011; **17:** 1627-1635.

161 Elson-Schwab I, Lorentzen A, Marshall CJ. MicroRNA-200 family members differentially regulate morphological plasticity and mode of melanoma cell invasion. *PLoS One* 2010; **5**.

162 Marasa BS, Srikantan S, Masuda K, Abdelmohsen K, Kuwano Y, Yang X *et al*. Increased MKK4 abundance with replicative senescence is linked to the joint reduction of multiple microRNAs. *Sci Signal* 2009; **2:** ra69.

163 Li X, Roslan S, Johnstone CN, Wright JA, Bracken CP, Anderson M *et al*. MiR-200 can repress breast cancer metastasis through ZEB1-independent but moesin-dependent pathways. *Oncogene* 2013.

164 Radhakrishnan P, Mohr AM, Grandgenett PM, Steele MM, Batra SK, Hollingsworth MA. MicroRNA-200c modulates the expression of MUC4 and MUC16 by directly targeting their coding sequences in human pancreatic cancer. *PLoS One* 2013; **8:** e73356.

165 Oishi N, Kumar MR, Roessler S, Ji J, Forgues M, Budhu A *et al*. Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma. *Hepatology* 2012; **56:** 1792-1803.

166 Cao H, Jheon A, Li X, Sun Z, Wang J, Florez S *et al*. The Pitx2:miR-200c/141:noggin pathway regulates Bmp signaling and ameloblast differentiation. *Development* 2013; **140:** 3348-3359.

167 Cama A, Verginelli F, Lotti LV, Napolitano F, Morgano A, D'Orazio A *et al*. Integrative genetic, epigenetic and pathological analysis of paraganglioma reveals complex dysregulation of NOTCH signaling. *Acta Neuropathol* 2013; **126:** 575-594.

168 Yang X, Ni W, Lei K. miR-200b suppresses cell growth, migration and invasion by targeting Notch1 in nasopharyngeal carcinoma. *Cell Physiol Biochem* 2013; **32:** 1288-1298.

169 Lerner M, Haneklaus M, Harada M, Grander D. MiR-200c regulates Noxa expression and sensitivity to proteasomal inhibitors. *PLoS One* 2012; **7:** e36490.

170 Howe EN, Cochrane DR, Cittelly DM, Richer JK. miR-200c targets a NF-kappaB up-regulated TrkB/NTF3 autocrine signaling loop to enhance anoikis sensitivity in triple negative breast cancer. *PLoS One* 2012; **7:** e49987.

171 Sun Y, Shen S, Liu X, Tang H, Wang Z, Yu Z *et al*. miR-429 inhibits cells growth and invasion and regulates EMT-related marker genes by targeting Onecut2 in colorectal carcinoma. *Mol Cell Biochem* 2014; **390:** 19-30.

172 Huang W, Liu H, Wang T, Zhang T, Kuang J, Luo Y *et al*. Tonicity-responsive microRNAs contribute to the maximal induction of osmoregulatory transcription factor OREBP in response to high-NaCl hypertonicity. *Nucleic Acids Res* 2011; **39:** 475-485.

173 Murray AR, Chen Q, Takahashi Y, Zhou KK, Park K, Ma JX. MicroRNA-200b downregulates oxidation resistance 1 (Oxr1) expression in the retina of type 1 diabetes model. *Invest Ophthalmol Vis Sci* 2013; **54:** 1689-1697.

174 Becker LE, Lu Z, Chen W, Xiong W, Kong M, Li Y. A systematic screen reveals MicroRNA clusters that significantly regulate four major signaling pathways. *PLoS One* 2012; **7:** e48474.

175 Zhang X, Zhang B, Gao J, Wang X, Liu Z. Regulation of the microRNA 200b (miRNA-200b) by transcriptional regulators PEA3 and ELK-1 protein affects expression of Pin1 protein to control anoikis. *J Biol Chem* 2013; **288:** 32742-32752.

176 Patel V, Hajarnis S, Williams D, Hunter R, Huynh D, Igarashi P. MicroRNAs Regulate Renal Tubule Maturation through Modulation of Pkd1. *J Am Soc Nephrol* 2012; **23:** 1941-1948.

177 Uhlmann S, Zhang JD, Schwager A, Mannsperger H, Riazalhosseini Y, Burmester S *et al*. miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene* 2010; **29:** 4297-4306.

178 Hu W, Wang X, Ding X, Li Y, Zhang X, Xie P *et al*. MicroRNA-141 represses HBV replication by targeting PPARA. *PLoS One* 2012; **7:** e34165.

179 Jurmeister S, Baumann M, Balwierz A, Keklikoglou I, Ward A, Uhlmann S *et al*. MicroRNA-200c represses migration and invasion of breast cancer cells by targeting actin-regulatory proteins FHOD1 and PPM1F. *Mol Cell Biol* 2012; **32:** 633-651.

180 Nagalla S, Shaw C, Kong X, Kondkar AA, Edelstein LC, Ma L *et al*. Platelet microRNA-mRNA coexpression profiles correlate with platelet reactivity. *Blood* 2011; **117:** 5189-5197.

181 Borbone E, De Rosa M, Siciliano D, Altucci L, Croce CM, Fusco A. Up-regulation of miR-146b and down-regulation of miR-200b contribute to the cytotoxic effect of histone deacetylase inhibitors on ras-transformed thyroid cells. *J Clin Endocrinol Metab* 2013; **98:** E1031-1040.

182 Li R, He JL, Chen XM, Long CL, Yang DH, Ding YB *et al*. MiR-200a is involved in proliferation and apoptosis in the human endometrial adenocarcinoma cell line HEC-1B by targeting the tumor suppressor PTEN. *Mol Biol Rep* 2014; **41:** 1977-1984.

183 Liao C, Chen W, Fan X, Jiang X, Qiu L, Chen C *et al*. MicroRNA-200c Inhibits Apoptosis in Pituitary Adenoma Cells by Targeting the PTEN/Akt Signaling Pathway. *Oncol Res* 2014; **21:** 129-136.

184 Shen LJ, He JL, Yang DH, Ding YB, Chen XM, Geng YQ *et al*. Mmu-microRNA-200a overexpression leads to implantation defect by targeting phosphatase and tensin homolog in mouse uterus. *Reprod Sci* 2013; **20:** 1518-1528.

185 You X, Liu F, Zhang T, Li Y, Ye L, Zhang X. Hepatitis B virus X protein upregulates oncogene Rab18 to result in the dysregulation of lipogenesis and proliferation of hepatoma cells. *Carcinogenesis* 2013; **34:** 1644-1652.

186 Ye F, Tang H, Liu Q, Xie X, Wu M, Liu X *et al*. miR-200b as a prognostic factor in breast cancer targets multiple members of RAB family. *J Transl Med* 2014; **12:** 17-27.

187 Liu Q, Tang H, Liu X, Liao Y, Li H, Zhao Z *et al*. miR-200b as a prognostic factor targets multiple members of RAB family in glioma. *Med Oncol* 2014; **31:** 859.

188 Li L, Tang J, Zhang B, Yang W, Liugao M, Wang R *et al*. Epigenetic modification of MiR-429 promotes liver tumour-initiating cell properties by targeting Rb binding protein 4. *Gut* 2014.

189 Xia W, Li J, Chen L, Huang B, Li S, Yang G *et al*. MicroRNA-200b regulates cyclin D1 expression and promotes S-phase entry by targeting RND3 in HeLa cells. *Mol Cell Biochem* 2010; **344:** 261-266.

190 Chang L, Guo F, Wang Y, Lv Y, Huo B, Wang L *et al*. MicroRNA-200c regulates the sensitivity of chemotherapy of gastric cancer SGC7901/DDP cells by directly targeting RhoE. *Pathol Oncol Res* 2014; **20:** 93-98.

191 Peng F, Jiang J, Yu Y, Tian R, Guo X, Li X *et al*. Direct targeting of SUZ12/ROCK2 by miR-200b/c inhibits cholangiocarcinoma tumourigenesis and metastasis. *Br J Cancer* 2013; **109:** 3092-3104.

192 Szczyrba J, Nolte E, Wach S, Kremmer E, Stohr R, Hartmann A *et al*. Downregulation of Sec23A protein by miRNA-375 in prostate carcinoma. *Mol Cancer Res* 2011; **9:** 791-800.

193 Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC *et al*. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc Natl Acad Sci U S A* 2010; **107:** 10514-10519.

194 Xiao J, Gong AY, Eischeid AN, Chen D, Deng C, Young CY *et al*. miR-141 modulates androgen receptor transcriptional activity in human prostate cancer cells through targeting the small heterodimer partner protein. *Prostate* 2012; **72:** 1514-1522.

195 Eades G, Yao Y, Yang M, Zhang Y, Chumsri S, Zhou Q. miR-200a regulates SIRT1 expression and epithelial to mesenchymal transition (EMT)-like transformation in mammary epithelial cells. *J Biol Chem* 2011; **286:** 25992-26002.

196 Baseler WA, Thapa D, Jagannathan R, Dabkowski ER, Croston TL, Hollander JM. miR-141 as a regulator of the mitochondrial phosphate carrier (Slc25a3) in the type 1 diabetic heart. *Am J Physiol Cell Physiol* 2012; **303:** C1244-1251.

197 Liu YN, Yin JJ, Abou-Kheir W, Hynes PG, Casey OM, Fang L *et al*. MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. *Oncogene* 2013; **32:** 296-306.

198 Braun J, Hoang-Vu C, Dralle H, Huttelmaier S. Downregulation of microRNAs directs the EMT and invasive potential of anaplastic thyroid carcinomas. *Oncogene* 2010; **29:** 4237-4244.

199 Shin JO, Lee JM, Cho KW, Kwak S, Kwon HJ, Lee MJ *et al*. MiR-200b is involved in Tgf-beta signaling to regulate mammalian palate development. *Histochem Cell Biol* 2012; **137:** 67-78.

200 Chen Y, Xiao Y, Ge W, Zhou K, Wen J, Yan W *et al*. miR-200b inhibits TGF-beta1-induced epithelial-mesenchymal transition and promotes growth of intestinal epithelial cells. *Cell Death Dis* 2013; **4:** e541.

201 Liao X, Xue H, Wang YC, Nazor KL, Guo S, Trivedi N *et al*. Matched miRNA and mRNA signatures from an hESC-based in vitro model of pancreatic differentiation reveal novel regulatory interactions. *J Cell Sci* 2013; **126:** 3848-3861.

202 Li J, Du L, Yang Y, Wang C, Liu H, Wang L *et al*. MiR-429 is an independent prognostic factor in colorectal cancer and exerts its anti-apoptotic function by targeting SOX2. *Cancer Lett* 2013; **329:** 84-90.

203 Buller B, Chopp M, Ueno Y, Zhang L, Zhang RL, Morris D *et al*. Regulation of serum response factor by miRNA-200 and miRNA-9 modulates oligodendrocyte progenitor cell differentiation. *Glia* 2012; **60:** 1906-1914.

204 Williams KC, Renthal NE, Condon JC, Gerard RD, Mendelson CR. MicroRNA-200a serves a key role in the decline of progesterone receptor function leading to term and preterm labor. *Proc Natl Acad Sci U S A* 2012; **109:** 7529-7534.

205 Cui Y, Chen J, He Z, Xiao Y. SUZ12 depletion suppresses the proliferation of gastric cancer cells. *Cell Physiol Biochem* 2013; **31:** 778-784.

206 Lin J, Liu C, Gao F, Mitchel RE, Zhao L, Yang Y *et al*. miR-200c enhances radiosensitivity of human breast cancer cells. *J Cell Biochem* 2013; **114:** 606-615.

207 Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S *et al*. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* 2008; **9:** 582-589.

208 Wang B, Koh P, Winbanks C, Coughlan MT, McClelland A, Watson A *et al*. miR-200a Prevents renal fibrogenesis through repression of TGF-beta2 expression. *Diabetes* 2011; **60:** 280-287.

209 Ali-Osman F, Rairkar A, Young P. Formation and repair of 1,3-bis-(2-chloroethyl)-1-nitrosourea and cisplatin induced total genomic DNA interstrand crosslinks in human glioma cells. *Cancer Biochem Biophys* 1995; **14:** 231-241.

210 Liu Y, Ding Y, Huang J, Wang S, Ni W, Guan J *et al*. MiR-141 Suppresses the Migration and Invasion of HCC Cells by Targeting Tiam1. *PLoS One* 2014; **9:** e88393.

211 Dai Y, Xia W, Song T, Su X, Li J, Li S *et al*. MicroRNA-200b is overexpressed in endometrial adenocarcinomas and enhances MMP2 activity by downregulating TIMP2 in human endometrial cancer cell line HEC-1A cells. *Nucleic Acid Ther* 2013; **23:** 29-34.

212 Xu L, Li Q, Xu D, Wang Q, An Y, Du Q *et al*. hsa-miR-141 downregulates TM4SF1 to inhibit pancreatic cancer cell invasion and migration. *Int J Oncol* 2014; **44:** 459-466.

213 Cochrane DR, Spoelstra NS, Howe EN, Nordeen SK, Richer JK. MicroRNA-200c mitigates invasiveness and restores sensitivity to microtubule-targeting chemotherapeutic agents. *Mol Cancer Ther* 2009; **8:** 1055-1066.

214 Prislei S, Martinelli E, Mariani M, Raspaglio G, Sieber S, Ferrandina G *et al*. MiR-200c and HuR in ovarian cancer. *BMC Cancer* 2013; **13:** 72.

215 Lee ST, Feng M, Wei Y, Li Z, Qiao Y, Guan P *et al*. Protein tyrosine phosphatase UBASH3B is overexpressed in triple-negative breast cancer and promotes invasion and metastasis. *Proc Natl Acad Sci U S A* 2013; **110:** 11121-11126.

216 Chang SH, Lu YC, Li X, Hsieh WY, Xiong Y, Ghosh M *et al*. Antagonistic function of the RNA-binding protein HuR and miR-200b in post-transcriptional regulation of vascular endothelial growth factor-A expression and angiogenesis. *J Biol Chem* 2012; **288:** 4908-4921.

217 McArthur K, Feng B, Wu Y, Chen S, Chakrabarti S. MicroRNA-200b regulates vascular endothelial growth factor-mediated alterations in diabetic retinopathy. *Diabetes* 2011; **60:** 1314-1323.

218 Sossey-Alaoui K, Bialkowska K, Plow EF. The miR200 family of microRNAs regulates WAVE3-dependent cancer cell invasion. *J Biol Chem* 2009; **284:** 33019-33029.

219 Tang H, Kong Y, Guo J, Tang Y, Xie X, Yang L *et al*. Diallyl disulfide suppresses proliferation and induces apoptosis in human gastric cancer through Wnt-1 signaling pathway by up-regulation of miR-200b and miR-22. *Cancer Lett* 2013; **340:** 72-81.

220 Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C *et al*. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 2007; **131:** 1109-1123.

221 Imanaka Y, Tsuchiya S, Sato F, Shimada Y, Shimizu K, Tsujimoto G. MicroRNA-141 confers resistance to cisplatin-induced apoptosis by targeting YAP1 in human esophageal squamous cell carcinoma. *J Hum Genet* 2011; **56:** 270-276.

222 Zhu ZM, Xu YF, Su QJ, Du JD, Tan XL, Tu YL *et al*. Prognostic significance of microRNA-141 expression and its tumor suppressor function in human pancreatic ductal adenocarcinoma. *Mol Cell Biochem* 2014; **388:** 39-49.

223 Capobianco V, Nardelli C, Ferrigno M, Iaffaldano L, Pilone V, Forestieri P *et al*. miRNA and protein expression profiles of visceral adipose tissue reveal miR-141/YWHAG and miR-520e/RAB11A as two potential miRNA/protein target pairs associated with severe obesity. *J Proteome Res* 2012; **11:** 3358-3369.

224 Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G *et al*. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 2008; **10:** 593-601.

225 Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F *et al*. miR-200 enhances mouse breast cancer cell colonization to form distant metastases. *PLoS One* 2009; **4:** e7181.

226 Yu KR, Lee S, Jung JW, Hong IS, Kim HS, Seo Y *et al*. MicroRNA-141-3p plays a role in human mesenchymal stem cell aging by directly targeting ZMPSTE24. *J Cell Sci* 2013; **126:** 5422-5431.

227 Bai WD, Ye XM, Zhang MY, Zhu HY, Xi WJ, Huang X *et al*. MiR-200c suppresses TGF-beta signaling and counteracts trastuzumab resistance and metastasis by targeting ZNF217 and ZEB1 in breast cancer. *Int J Cancer* 2014; **135:** 1356-1368.

228 Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* 2008; **283:** 14910-14914.

229 Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D *et al*. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* 2009; **138:** 592-603.

230 Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A *et al*. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 2010; **7:** 64-77.

231 Feng B, Wang R, Song HZ, Chen LB. MicroRNA-200b reverses chemoresistance of docetaxel-resistant human lung adenocarcinoma cells by targeting E2F3. *Cancer* 2012; **118:** 3365-3376.

232 Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F *et al*. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* 2013; **154:** 61-74.

233 Gravgaard KH, Lyng MB, Laenkholm AV, Sokilde R, Nielsen BS, Litman T *et al*. The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. *Breast Cancer Res Treat* 2012; **134:** 207-217.

234 Dai Y, Xia W, Song T, Su X, Li J, Li S *et al*. MicroRNA-200b Is Overexpressed in Endometrial Adenocarcinomas and Enhances MMP2 Activity by Downregulating TIMP2 in Human Endometrial Cancer Cell Line HEC-1A Cells. *Nucleic Acid Ther* 2013.

235 Pichler M, Ress AL, Winter E, Stiegelbauer V, Karbiener M, Schwarzenbacher D *et al*. MiR-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. *Br J Cancer* 2014; **110:** 1614-1621.

236 Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M *et al*. Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. *Cancer Res* 2010; **70:** 5226-5237.

237 Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL *et al*. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 2008; **105:** 10513-10518.

238 Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol* 2008; **110:** 13-21.

239 Toiyama Y, Hur K, Tanaka K, Inoue Y, Kusunoki M, Boland CR *et al*. Serum miR-200c Is a Novel Prognostic and Metastasis-Predictive Biomarker in Patients With Colorectal Cancer. *Ann Surg* 2014; **259:** 735-743.

Chapter 2

The role of the miR-200 family in EMT

This chapter was adapted from the following publication:

Perdigão-Henriques R, Petrocca F, Altschuler G, Thomas MP, Le MT, Tan SM, Hide W, Lieberman J. miR-200 promotes the mesenchymal to epithelial transition by suppressing multiple members of the Zeb2 and Snail1 transcriptional repressor complexes. *Oncogene*. Advanced online publication Mar 23, 2015. doi: 10.1038/onc.2015.69

Summary

The miR-200 family promotes the epithelial state by suppressing the Zeb1/Zeb2 epithelial gene transcriptional repressors. To identify other miR-200-regulated genes, we isolated mRNAs bound to transfected biotinylated miR-200c in mouse breast cancer cells. 520 mRNAs were significantly enriched in miR-200c binding at least 2-fold. Putative miR-200-regulated genes included *Zeb2*, enriched 3.5-fold in the pull-down. However, *Zeb2* knockdown does not fully recapitulate miR-200c over-expression, suggesting that regulating other miR-200 targets contributes to miR-200's enhancement of epithelial gene expression. Candidate genes were highly enriched for miR-200c seed pairing in their 3'UTR and CDS and for genes that were down-regulated by miR-200c over-expression. EGFR and downstream MAPK signaling pathways were the most enriched pathways. Genes whose products mediate TGF-β signaling were also significantly over-represented, and miR-200 counteracted the suppressive effects of TGF-β and BMP-2 on epithelial gene expression. miR-200c regulated the 3'UTRs of 12 of 14 putative miR-200c-binding mRNAs tested. The extent of mRNA binding to miR-200c strongly correlated with gene suppression. Twelve targets of miR-200c (*Crtap, Fhod1, Smad2, Map3k1, Tob1, Ywhag/14-3-3γ, Ywhab/14-3-3β, Smad5, Zfp36, Xbp1, Mapk12*, *Snail1*) were experimentally validated by identifying their 3'UTR miR-200 recognition elements. Smad2 and Smad5 form a complex with Zeb2 and Ywhab/14-3-3β and Ywhag/14-3-3γ form a complex with Snail1. These complexes that repress transcription assemble on epithelial gene promoters. miR-200 over-expression induced RNA polymerase II localization and reduced Zeb2 and Snail1 binding to epithelial gene promoters. Expression of miR-200-resistant *Smad5* modestly, but significantly, reduced epithelial gene induction by miR-200. miR-200 expression and *Zeb2* knockdown are known to inhibit cell invasion in *in vitro* assays. Knockdown of each of 3 novel miR-200 target genes identified

here, *Smad5*, *Ywhag* and *Crtap*, also profoundly suppressed cell invasion. Thus miR-200 suppresses TGF-β/BMP signaling, promotes epithelial gene expression and suppresses cell invasion by regulating a network of genes.

CONTENTS

1. Introduction

The miR-200 family promotes the epithelial state of cells and suppresses mesenchymal properties during development and cellular differentiation. Transitions between epithelial and mesenchymal states (epithelial to mesenchymal (EMT) and mesenchymal to epithelial (MET)), which alter the motility of cells during development[1], play an important role in tumor metastasis[2]. Epithelial cells express high levels of E-cadherin, occludin and claudins, while mesenchymal cells express high levels of N-cadherin, vimentin and fibronectin[3]. Some tumor cell lines, especially less differentiated cells, show plasticity in their display of epithelial and mesenchymal traits. Although the early steps of metastases - penetration through the basement membrane, invasion of the vasculature and extravasation - are facilitated by cells acquiring mesenchymal traits, the ability to colonize distant tissues and form macroscopic metastases may be facilitated by epithelial properties.[1,4-6] Accordingly, miR-200 expression has been associated with both increased[4,7-10] and decreased[11-13] malignancy and metastases, depending on the cancer model. High levels of circulating miR-200 are associated with poor prognosis in several human cancers, including ovarian, prostate, pancreatic and metastatic colorectal cancers[14-17].

The miR-200 miRNAs are five homologous miRNAs (miR-141, miR-429, miR-200a, miR-200b and miR-200c), whose sequences are conserved

between human and mouse. These miRNAs belong to two families whose seed regions (nucleotides 2 to 7 of the mature active strand, which form the most important region for target recognition) differ by a single nucleotide. The seed for miR-200b/c/429 sequence is 5'-AAUACU-3', while for miR-200a/141 it is 5'-AACACU-3'. Because of its role as a master regulator of MET, miR-200 has been extensively studied. The Zeb epithelial gene transcriptional repressors contain multiple miR-200 binding sites in their 3'UTR (*Zeb1* has 5 and *Zeb2* has 6 conserved predicted binding sites) and their expression is profoundly downregulated by miR-200. Much of miR-200's effect on increasing epithelial gene expression and inhibiting cell motility and invasivity can be reproduced by knocking down *Zeb1* and/or *Zeb2.*[4,18-21] Many other gene targets have been described in mouse or man (Table S1), including the transcription factors *Ets1*[22], *Snail1* (*Snail*)[23] and *Snail2* (*Slug*)[24]. miR-200 has also been shown to regulate cell-cell adhesion[18-19,21], cancer stem cell self-renewal and differentiation[25-27], cell division and apoptosis[28-31] and chemoresistance.[32-33]

miRNAs accomplish their biological functions by regulating networks of genes[34]. Transfection of a biotinylated (Bi-)miRNA mimic and streptavidin pull-down (PD) of associated mRNAs provides an unbiased method to identify the genes regulated by a particular miRNA with high specificity[34-39]. To better understand how miR-200 functions, we carried out a Bi-miR-200c PD in the mouse breast cancer cell line 4TO7. 4TO7 is a mouse triple

negative breast cancer (TNBC) cell line that is able to perform all the steps of metastasis when implanted in the mammary fat pad, except colonization of metastatic sites. 4TO7 does not express the miR-200 family, but ectopic expression of miR-200c, which accounts for 93% of mature miR-200b/c/429 family expression in human breast tumors[40], leads to efficient colonization of lung and liver, macroscopic metastases and reduced survival.[4,8] Moreover, 4T1E, an isogenic clone derived from the same primary, spontaneously arising tumor that metastasizes, differs from 4TO7 by high expression of the miR-200 family. mRNAs for 520 genes, including *Zeb2* and other known miR-200 target genes, were significantly enriched at least 2-fold in the PD with miR-200c relative to a *C. elegans* control miRNA. The pulled down genes were highly enriched for miR-200c seed-binding sequences and for genes knocked down by miR-200 over-expression. Of 18 genes chosen for experimental validation, 14 (78%) were confirmed as significantly enriched by PD and qRT-PCR. Of those, 12 (86%) had functional miR-200 recognition elements in their 3'UTR. Genes that participate in EGFR and TGF-β/BMP signaling were highly over-represented in the miR-200 target gene set. In addition to *Zeb2* and *Snail1*, we found that miR-200 also suppressed expression of their cofactors, *Smad2/Smad5* and *Ywhab/Ywhag*, respectively.[41-42] Suppression of these multiple epithelial gene regulators enhanced the ability of miR-200 to promote MET. *Crtap*, which is required for type I collagen prolyl-hydroxylation, was a novel strong target of miR-200, whose knockdown

strongly inhibited invasivity through Matrigel. Crtap is postulated to facilitate TGF-β binding to extracellular matrix.

2. Materials and methods

2.1. Cell culture

4TO7 and 4T1 cells were kindly provided by Fred Miller (Wayne State University).[43] 4T1E cells were previously generated in our laboratory[44]. 4TO7, 4T1, 4T1E and NIH/3T3 cell lines were grown in DMEM (Gibco) supplemented with 10% fetal bovine serum (FBS), 1 mM L-glutamine and penicillin/streptomycin (Gibco).

2.2. siRNA and miRNA mimics transfection

4TO7, 4T1 or NIH/3T3 cells were transfected with 50 nM negative control miRNA mimic (Dharmacon, CN-001000-01-10), miR-200c miRNA mimic (Dharmacon, C-310569-07-0010), control siRNA (Dharmacon, D-001210-04-05) or siRNAs against *Zeb2* (Dharmacon, M-059671-01-0005), *Snail1* (Dharmacon, M-062765-00-0005), *Ywhag* (Dharmacon, M-059307-00-0005), *Smad5* (Dharmacon, M-057015-00-0005), *Crtap* (Dharmacon, M-049986-00-0005) or Firefly luciferase (sequence listed in Table S6) using Lipofectamine 2000 (Invitrogen). Cells were incubated with miRNA or siRNA-containing lipid complexes, according to the manufacturer's protocol, for 4 h before culture supernatants were removed and replaced

with complete growth medium. Cells were harvested 48 or 72 h post transfection for mRNA and protein analyses.

2.3. RNA isolation and quantitative RT–PCR

Total RNA was isolated using Trizol reagent (Ambion), treated with DNase I (Promega) and reverse transcribed using random hexamers and Thermoscript RT kit (Invitrogen) as per the manufacturer's protocols. Quantitative real time PCR (qRT-PCR) was performed in triplicate samples using Ssofast^{TM} Evagreen qPCR assay (BioRad) on a BioRad CFX96 C1000 Thermal Cycler. mRNA levels were normalized to the housekeeping gene *Gapdh*. miRNA was quantified in triplicate using the TaqMan MicroRNA Assay (Applied Biosystems) as per the manufacturer's instructions and normalized to snoRNA234. Primer sequences are listed in Table S6.

2.4. Immunoblot

Whole cell lysates were prepared using RIPA buffer (150 nM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris pH 8.0) supplemented with Complete Mini-protease Inhibitor Cocktail (Roche). For phosphoprotein analysis, lysis buffer was further supplemented with 10 mM NaF, 2 mM Na_3VO_4 and 2.5 mM sodium pyrophosphate. Protein concentration was determined using the BCA Protein Assay (Pierce). Samples were resolved on SDS-Page gels and transferred using a Transblot semi-dry transfer apparatus (BioRad). Blots were probed with

antibodies to Smad5 (sc-7443) and Ywhab (sc-59419) from Santa Cruz Biotechnology; Ywhag (EMD Millipore, 05-639); Claudin-3 (Life Technologies, 341700); Crtap (kind gift of Prof. Brendan Lee, Baylor College of Medicine); Snail1 (3895), phospho-Smad1/5/8 (9511), Smad2 (5339) and phospho-Smad2 (3101) from Cell Signaling; E-cadherin (BD Biosciences, 61081); and α-Tubulin (Sigma, T5168). Blots were quantified by densitometry.

2.5. Bi-miR-200c pull-down

4TO7 cells (2.4×10^6) were transfected as described above with 50 nM Bi-miR-200c or Bi-cel-miR-67 miRNA mimics biotinylated at the 3'-end of the active strand (Dharmacon). Biotin PDs were performed 24 h after transfection as previously described[34]. The level of mRNA in the Bi-miR-200c or Bi-cel-miR-67 control PD was then quantified by mRNA microarray (Illumina) and analyzed as described below. For technical validation of the PD using qRT-PCR, cells were transfected as described above and an optimized version of the PD protocol was used[39]. In this case, mRNA levels in the PD and input samples were first normalized to *Gapdh* and then the enrichment ratio of the control-normalized PD mRNA to the control-normalized input levels was calculated.

2.6. Microarray analysis

Total RNA (from two independent experiments) was amplified, labeled and hybridized to Illumina MouseRef-8 v2.0 expression beadchip (Illumina)

mouse whole-genome expression arrays. The quality of the RNA was assessed before performing the microarray and the quality of the microarray data was assessed using affyPLM and Affy software. The microarray data were normalized using RMA[45] to reduce interarray variation. Probes were mapped to Entrez Gene identifications (IDs) using systematically updated annotations from AILUN (Array Information Library Universal Navigator)[46]. The PD enrichment ratio was defined as (Bi-miR-200c PD/Bi-cel-miR-67 PD)/(Bi-miR-200c input/Bi-cel-miR-67 input) and calculated for each probe. A detection *P* cutoff of 0.05 was used to exclude probes in any of the control arrays. For each probe, a significance value associated with the enrichment ratio was obtained using a *t*-test that compared the PD to input ratio for the Bi-miR-200c experiments to the control Bi-cel-miR-67 experiments. Probes with a *P* of less than 0.01, and an enrichment ratio of greater than 2 were considered 'hits' for the informatic analysis of the PD data.

2.5. Bi-miR-200c pull-down

The fold change in probe expression caused by Bi-miR-200c over-expression was calculated by the ratio of input RNA from Bi-miR-200c-transfected 4TO7 cells to the ratio of input RNA from Bi-cel-miR-67-transfected cells. A cumulative distribution function plot was used to compare the distribution of Bi-miR-200c over-expression fold changes for

PD gene lists defined by different cutoffs. The Kolmogorov-Smirnov [KS] test was used for statistical comparisons between gene sets.

2.8. Analysis of miR-200c target genes by target prediction algorithms

To determine whether a gene was also a predicted target of miR-200c, the presence of miR-200c binding sites was analyzed using TargetScan 6.2 (www.targetscan.org/)[47-49], PITA (genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html)[50] or RNAhybrid (bibiserv.techfak.uni-bielefeld.de/rnahybrid)[51].

2.9. Hexamer analysis

The mature mmu-miR-200c sequence was downloaded from miRBase; all RefSeq mouse mRNAs were obtained from NCBI. Hexamers complementary to each position of the miR-200c sequence were found in the complete gene set and their frequency was length normalized (matches/kb). Significance of enrichment was calculated by Monte Carlo sampling without replacement from the complete expressed gene set as described.[34]

2.10. Pathway enrichment analysis and network visualization

Canonical pathway gene sets were compiled from Wikipathways[52] and genes not expressed in 4TO7 cells (as determined by using Bi-cel-miR-67-transfected 4TO7 microarray data using a detection *P* cutoff of 0.05) were

removed. The hypergeometric test was used to assess the enrichment of these biological gene sets in the list of Bi-miR-200c PD genes. The interactome of directly-interacting pulled-down genes was generated using the Ingenuity Pathway Analysis software (Ingenuity Systems, www.ingenuity.com). Edges represent direct functional interactions (protein-protein or protein-DNA/RNA interactions) between gene products.

2.11. miRNA and mRNA correlation in NCI-60 cell lines and primary human breast cancers

The relationship between *CRTAP*, *SDHA*, *ZEB2* and miR-200c expression levels in the NCI-60 panel of cancer cell lines was analyzed from genome-wide mRNA and miRNA profiling data using Cellminer (http://discover.nci.nih.gov/cellminer)[53]. Matched miRNA and mRNA profiling in human primary breast tumor samples was reported in GSE19783 deposited in GEO (www.ncbi.nlm.nih.gov/geo/). Multiple probes mapping to the same gene were aggregated using the median expression. Pearson correlation with a one-sided test for negative correlation was used to investigate the correlation between miR-200c and *CRTAP*, *ZEB2* and *SDHA* expression levels.

2.12. Dual-luciferase reporter assay

The full 3'UTR of each gene was amplified by PCR from genomic DNA isolated from 4TO7 cells and cloned into the psiCHECK-2 vector (Promega) immediately upstream of the *Renilla* luciferase gene. These constructs

were used to generate constructs containing mutations in putative miR-200c MREs within the full 3'UTR of indicated genes. The sequences of all constructs were verified by sequencing. 4T1E cells were cotransfected with 50 nM miRNA mimics together with psiCHECK-2 vector containing the 3'UTR of indicated genes. Cells were lysed 24 h after transfection in Passive Lysis Buffer (Promega) and Firefly and *Renilla* luciferase activities were measured using the Dual Luciferase Assay System (Promega) on a Synergy2 plate reader (Biotek). *Renilla* luciferase activity levels were normalized to Firefly luciferase and results from three independent experiments were compared to activity in cells expressing empty psiCHECK-2 vector. Sequence of primers used for cloning 3'UTRs and to generate mutated 3'UTRs are listed in Table S6.

2.13. E-cadherin promoter reporter assay

The -178/+92 fragment of the mouse E-cadherin promoter was amplified by PCR from genomic DNA isolated from 4TO7 cells using oligonucleotides linked to KpnI and XhoI restriction sites and cloned into the same restriction sites in the pGL3 vector (Invitrogen) fused to the Firefly luciferase reporter gene. To determine the activity of the E-cadherin promoter, 4TO7 cells (2×10^4 cells) were co-transfected in 24-well plates with 50 ng of the reporter and 50 ng of TK-*Renilla* plasmid (Promega). Cells were transfected with 50 nM siRNA or miRNA 24 h later and 48 h after that cells were then lysed in Passive Lysis Buffer (Promega) and Firefly and *Renilla* luciferase activities

were measured consecutively using the Dual Luciferase Assay System (Promega) on a Synergy2 plate reader (Biotek). Firefly luciferase activity was normalized to *Renilla* luciferase activity and relative to activity measured for control siRNA or miRNA transfection.

2.14. Chromatin immunoprecipitation assay (ChIP)

4TO7 cells (12.5×10^6) were transfected with 50 nM miRNA as above and 24 h later cross-linked with 1% formaldehyde for 10 min at room temperature. Crosslinking was stopped by adding glycine to a final concentration of 0.125 M. Cells were washed with cold PBS and kept on ice for 10 min in 25 mM HEPES (pH 7.8), 1.5 mM $MgCl_2$, 10 mM KCl, 0.1% NP-40, 1 mM DTT and Complete Mini-protease Inhibitor Cocktail (Roche). Following homogenization, nuclei were pelleted and resuspended in sonication buffer (50 mM HEPES (pH 7.9), 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate, 0.1% SDS, and Complete Mini-protease Inhibitor Cocktail (Roche)) and sonicated on ice to an average length of 200–1500 bp. Chromatin from ~5×10^6 cells was used for each immunoprecipitation, which was carried out overnight in the presence of 25 µL Protein G Dynabeads (Life Technologies, 10004D) using 2 µg of mouse monoclonal antibody against RNA Polymerase II (Santa Cruz Biotechnology, sc-56767), 5 µg of rabbit antiserum against Zeb2 (kind gift of Prof. Danny Huylebroeck, University of Leuven), 1 µg of rabbit antiserum against Snail1+Snail2 (Abcam, ab85931; we confirmed by qRT-PCR (data

not shown) that *Snail2* is not expressed in 4TO7 cells) or equal amounts of rabbit IgG (Cell Signaling, 2729S) or mouse IgG (Santa Cruz, sc-2025). Samples were then extensively washed with sonication buffer, wash buffer A (50 mM Hepes pH 7.9, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate, 0.1% SDS and Complete Mini-protease Inhibitor Cocktail (Roche)), wash buffer B (20 mM Tris, pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, 0.5% Na deoxycholate and Complete Mini-protease Inhibitor Cocktail (Roche)) and Tris-EDTA buffer to remove unbound DNA. Immunoprecipitated DNA was then eluted using elution buffer (50 mM Tris, pH 8.0, 1 mM EDTA, 1% SDS, 50 mM $NaHCO_3$) and treated with proteinase K (Ambion, AM2546) and RNAse (Sigma, R4642) overnight at 65ºC. DNA was then purified using a PCR purification kit (Qiagen) and analyzed by qRT-PCR using the primers listed in Table S6.

2.15. Transwell invasion assays

4TO7 cells (4×10^5) were transfected in 6-well plates with 50 nM siRNA or miRNA mimics as described above and 48 h later were incubated in serum-free medium for 2 h. Cells were then harvested with TrypLE™ Express (Invitrogen) and counted using a hemocytometer. A BD BioCoat Tumor Invasion System (BD Biosciences, 354165) containing BD Falcon Fluoroblock 24-Multiwell inserts (8-µm pore size; PET membrane) was used to measure invasion following the manufacturer's instructions. Briefly, 1.5 x 10^5 cells were added in serum-free medium to each insert and

complete medium containing 10% FBS was added into each basal chamber. After 20 h, invaded cells were quantified by labeling with calcein AM (BD Biosciences), and measuring the fluorescence underneath the membrane using a Synergy2 plate reader (Biotek).

2.16. Smad5 and Ywhag expression plasmids

Smad5 and *Ywhag* CDS were cloned into the multiple cloning site of pcDNA3.1(+) (Invitrogen) by RT-PCR using mRNA from 4TO7 cells. Primers used for PCR amplification are listed in Table S6.

2.17. TGF-β treatment

4TO7 cells (4×10^5) were transfected in 6-well plates with 50 nM miRNA mimics as described above. Culture supernatants were removed after 4 h and replaced with complete medium containing 5 ng/mL TGF-β (Cell Signaling, 5231) or the same volume of vehicle (20 mM sodium citrate, pH 3.0). After 24 h, medium was replaced with fresh medium containing TGF-β or vehicle. Cells were harvested 24 h after that for mRNA analysis.

2.18. Plasmid DNA transfection and BMP-2 treatment

4TO7 cells (4×10^5) were co-transfected in 6-well plates as described above with 6 μg empty vector, *Ywhag* or *Smad5* expression vectors together with 50 nM miRNA mimics. Culture supernatants were removed after 4 h and replaced with complete medium containing 25 ng/mL BMP-2 (Cell Signaling, 4697) or the same volume of vehicle (10 mM HCl, 0.1% BSA).

After 24 h, medium was replaced with fresh medium containing BMP-2 or vehicle. Immunoblot was performed on cell lysates harvested 24 h after that.

2.19. Cell viability assay

4TO7 cells (2×10^4), transfected in 24-well plates with 50 nM siRNA or miRNA mimics, were analyzed 48 h later for cell viability by CellTiter-Glo assay (Promega) following the manufacturer's protocol. An siRNA targeting *Plk1* (Dharmacon, M-040566-01-0005) was used as a control. Signals were normalized to control siRNA or miRNA treatment.

2.20. E-cadherin immunostaining

4TO7 cells, transfected with siRNA or miRNA mimics as described above, were plated 24 h later (10^4 cells/well) on a Lab-Tek chamber slide (Thermo Fisher, 177445) in 0.2 mL and fixed 48 h later in 4% formaldehyde in PBS. The slides were washed twice with PBS, then incubated in blocking buffer (PBS with 0.5% BSA and 0.05% Triton-X) for 30 min, then for 1 h in blocking buffer plus E-cadherin antibody (BD Biosciences, 61081) at a final dilution of 1:100. After washing, slides were placed in blocking buffer with Alexa-Fluor-488-conjugated donkey anti-mouse secondary antibody (Life Technologies, A-21202) for 1 h and then washed and stained with 4 μg/mL DAPI in PBS and mounted using polyvinyl alcohol (Sigma, P-8136) aqueous mounting medium. Cells were imaged using an Axiovert 200M

microscope (Pan Apochromat, 1.4 NA; Carl Zeiss) and analyzed with SlideBook 4.2 (Intelligent Imaging Innovations Inc.).

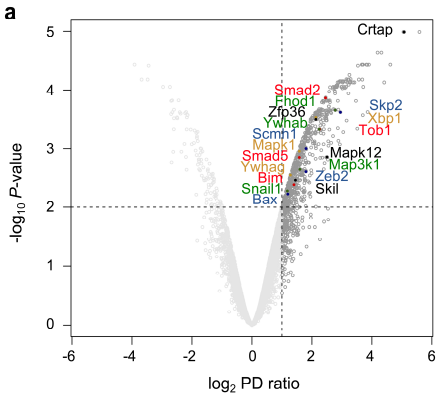
2.21. Statistical analysis

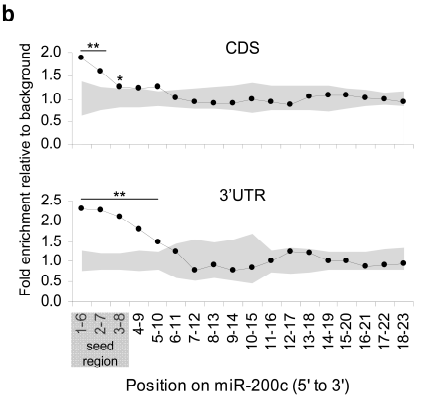
In all experiments, except where indicated, paired Student's *t*-tests (two-tailed) were used to analyze the difference between two groups of data.

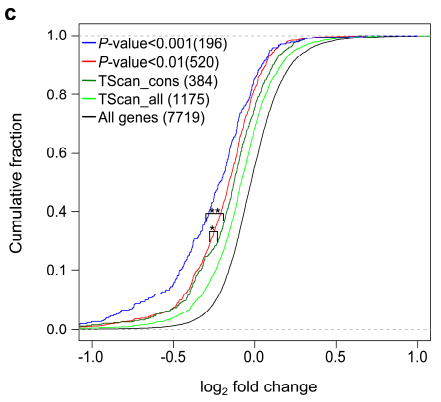
3. Results

3.1. miR-200c pull-down

A previously published protocol[34], which showed 90% specificity for identifying miR-34a-regulated genes, was used to capture miR-200-bound mRNAs. 4TO7 cells were transfected with control *C. elegans* miRNA mimic (Bi-cel-miR-67) or miR-200c mimic (Bi-miR-200c), both biotinylated at the 3'-end of the mature active strand. The next day cells were lysed and streptavidin-coated beads were used to isolate mRNAs bound to the biotinylated miRNAs. mRNA abundance in Bi-cel-miR-67 or Bi-miR-200c PD and input samples from 2 independent experiments were assessed on mRNA microarrays. An enrichment ratio R ((Bi-miR-200c PD/Bi-cel-miR-67 PD)/(Bi-miR-200c input/Bi-cel-miR-67 input)) was calculated as a measure of how much each transcript was enriched in the PD relative to its expression in the miRNA-transfected cells. We then calculated the average enrichment ratio and associated *P*-value for each microarray probe. 520 transcripts had an enrichment ratio >2 and $P<0.01$ (Figure 11a, Table S2). To determine the sensitivity of the Bi-miR-200c PD, we first determined how many of the known mouse miR-200 family published targets were enriched (Table S1). 13 experimentally validated mouse targets of any miR-200 family member are expressed in 4TO7 cells (*Zeb1* and *Snail2* are not expressed in these cells) and 9 of these are known to be regulated by miR-200c. 7 of the 13 (54%) and 5 of the 9 (56%) were identified in the PD.







d

	AdjP	nGenes (nPathway)
MAPK signaling pathways	7.9E-07	17 (162)
EGFR1 Signaling Pathway	7.9E-07	17 (167)
B Cell Receptor Signaling Pathway	8.9E-05	13 (145)
T Cell Receptor Signaling Pathway	1.2E-03	10 (118)
Signal Transduction of S1P Receptor	1.8E-03	4 (21)
Cholesterol Biosynthesis	2.1E-03	3 (12)
Integrin-mediated cell adhesion	2.1E-03	8 (90)
TGF-beta signaling pathways	4.5E-03	11 (172)
Kit Receptor Signaling Pathway	5.8E-03	6 (64)
Signaling of HGF Receptor	6.2E-03	4 (31)
IL-3 Signaling Pathway	9.9E-03	7 (95)
Sphingolipid Metabolism	9.9E-03	4 (36)
Urea cycle and metabolism of amino groups	9.9E-03	3 (21)

Figure 11. Bi-miR-200c PD identifies 520 putative miR-200c target genes (a) Volcano plot for Bi-miR-200c PD showing the log2-transformed enrichment ratio for each probe detected above background in the microarray versus its corresponding log10-transformed *P*. Probes with a PD-enrichment ratio R>2 are shown in dark gray and the horizontal dashed line corresponds to *P*=0.01. 520 genes fit the criteria R>2 and *P*<0.01. Probes detecting *Zeb2* (a known miR-200 target) and each of the 18 putative miR-200c targets chosen for validation are labeled (if there were multiple probes for the same gene, the one with the highest R was labeled). (b) Fold enrichment of hexamer sequences complementary to each 6-nucleotide segment of mature miR-200c in the CDS or 3'UTR of the 520 PD genes. Grey shading indicates hexamer enrichment ratios that did not differ by *P*<0.01 from that in equal-sized gene sets drawn randomly from all expressed genes. *, *P*=0.002; **, *P*<0.0001. (c) Cumulative distribution plots of mRNA reduction in 4TO7 cells after transfection with miR-200c vs cel-miR-67 comparing the set of genes enriched in the Bi-miR-200c PD (R>2 and *P*<0.01 (red) or *P*<0.001 (blue)) with all expressed genes detected above background in the microarray (black), and TargetScan predictions of evolutionarily conserved (dark green) or all (light green) miR-200c target genes. Numbers in parentheses indicate the number of genes in each set. (*, Kolmogorov–Smirnov test *P*=0.02; **, *P*=1.1E-6). (d) Canonical pathways (Wikipathways) most enriched (adjusted *P* (Adj*P*)<0.01) in the Bi-miR-200c PD (R>2, *P*<0.01). The PD genes in each enriched canonical pathway are listed in Table S3.

To estimate the false-positive rate of the Bi-miR-200c PD, we counted the number of transcripts that specifically bound to the control worm Bi-cel-miR-67, which has no mammalian orthologue, using the same criteria (enrichment ratio (Bi-cel-miR-67 PD/Bi-miR-200c PD)/(Bi-cel-miR-67 input/Bi-miR-200c input) >2 and P<0.01). Comparing 81 mRNAs enriched for binding to Bi-cel-miR-67 as background with 520 transcripts enriched for binding to Bi-miR-200c, the estimated false-positive rate is 15.6%.

The putative miRNA target genes had properties expected of canonical miR-200c-regulated genes. 43% of the 520 enriched mRNAs were predicted miR-200c targets by TargetScan. These therefore contain a perfect match to the miR-200c seed in their 3'UTR. The CDS and 3'UTRs (Figure 11b), but not the 5'UTRs (data not shown), of the 520 genes were also significantly enriched for hexamer seed matches (P<0.0001). Transcripts of the 520 pulled-down genes compared to all genes expressed in 4TO7 were significantly downregulated after Bi-miR-200c transfection (P<2.2E-16, Figure 11c). The extent of their downregulation was significantly greater than that of the 384 evolutionarily conserved, TargetScan predicted target genes (P=0.023) or the 1175 TargetScan predicted genes that included nonconserved binding sequences (P=2.1E-15). 196 genes were enriched in the PD using a more stringent P cut-off of <0.001. This gene set was even more strongly down-regulated than the

larger set of 520 genes enriched by *P*<0.01 (*P*=1.8E-3) or the conserved TargetScan target gene set (*P*=1.1E-6).

3.2. miR-200 bound genes are enriched for genes involved in EGFR, MAPK and TGF-β signaling and metabolic pathways implicated in metastasis and oncogenic transformation

To understand better the biological processes regulated by miR-200, we first performed pathway enrichment analysis of the 520 enriched PD genes (Figure 11d, Table S3). The most enriched pathways were MAPK and EGFR signaling, both with *P*-values of 8E-07, each containing 17 putative miR-200c targets, of which 6 overlapped. The MAPK and EGFR signaling were also significantly enriched pathways in the more stringently chosen gene list of 196 putative targets (R>2, *P*<0.001) (Figure S1a, Table S4). Other cell surface receptor signaling pathways were also enriched, such as the T and B cell receptor and Kit signaling pathways, which overlap with the EGFR pathway and use MAPK signaling. However, these other receptors are not expressed in breast tumor cells. TGF-β signaling genes were also significantly overrepresented (*P*<0.005) with 11 annotated genes. These included the *Zeb2*, *Smad2*, *Smad5*, and *Tgif* transcription factors that help control epithelial gene expression (see below) and other oncogenic transcription factors, including *Fosb* and *Skil*. The sphingosine-1-phosphate receptor signaling pathway, implicated in tumor metastasis, as well as enzymes in sphingolipid and cholesterol biosynthesis pathways, were also overrepresented. The mRNAs for 3 enzymes involved in amino acid

metabolism, which is altered in cancer cells, were also significantly enriched.

3.3. miR-200-bound genes form a dense interaction network

The 520 putative miR-200c target genes were analyzed by Ingenuity Pathway Analysis software (Ingenuity Systems, www.ingenuity.com) to identify directly-interacting target gene products. 222 gene products enriched in the PD were annotated to interact directly by protein-protein or protein-DNA/RNA interactions (Figure 12). Of note, the interactome contained a small module composed of genes that regulate transcriptional changes during EMT. These included the master epithelial gene suppressors *Snail1* and *Zeb2*, the *Snail1* co-repressors *Ywhag* (*14-3-3γ*) and *Ywhab* (*14-3-3β*)[42], the *Zeb2* co-repressors *Smad2* and *Smad5*[41] and Smad corepressor*, Tgif1*. Relieving suppression of epithelial gene expression may also be enhanced by inhibition of *Anp32e*, which removes histone H2A.Z from promoters and enhancers to initiate transcription from paused RNA pol II[54]. *Snail1*, *Zeb2*, *Smad2* and *Smad5* are activated by the canonical TGF-β and BMP signaling pathways, suggesting that miR-200's induction of MET is accomplished in part by suppressing TGF-β and BMP signaling. *Map2k6* and *Mapk12* in the direct-interaction network also mediate Smad-independent TGF-β responses.

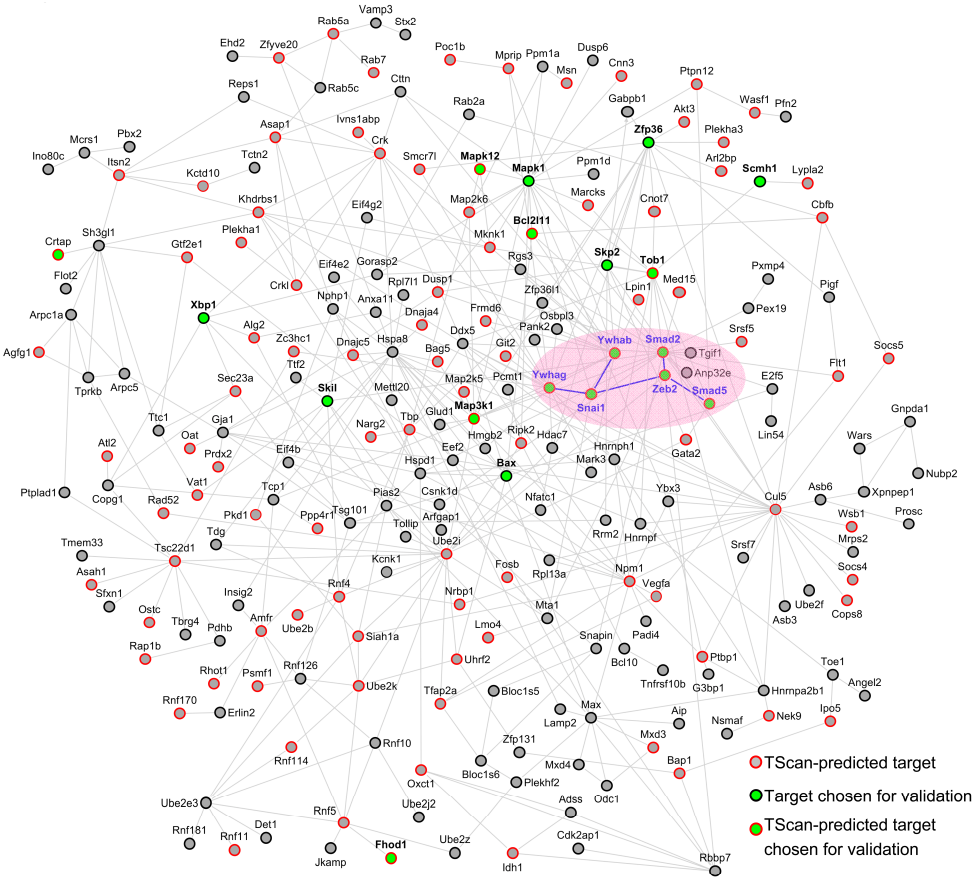


Figure 12. Interactome of directly-interacting pulled-down genes. 222 gene products enriched in the PD were annotated to interact directly. 45% of them (red border) are predicted by TargetScan (TScan) to be miR-200c targets in mice. 18 genes (green) were selected for experimental validation. Edges represent direct protein-protein or protein-DNA/RNA interactions. A module of genes involved in regulating transcription of epithelial genes is highlighted in pink.

These results suggest that miR-200 inhibits both Smad-dependent and -independent TGF-β and BMP signaling pathways. Another EMT-promoting transcription factor is *Xbp1*, which activates EMT in response to endoplasmic reticulum stress[55] and has a major role in the tumorigenicity, progression and recurrence of TNBC.[56] Genes known to regulate cell

adhesion such as *Fhod1*[57], *Wave1/Wasf1*[58] and *Map3k1*[59] were also in the miR-200 target gene interactome.

3.4. Experimental validation of the predictions of the Bi-miR-200c pull-down

To determine how reliable the predictions of miR-200-regulated genes were, we selected 18 putative target genes to test individually. These genes, which covered the full range of enrichment in the PD (2.3-33.7 fold, Figure 11a), were chosen because they were located at hubs in the interactome, were exceptionally enriched in the PD or might be related to miR-200's known functions. In particular, we were particularly interested in the five genes (*Ywhag*, *Ywhab*, *Snail1*, *Smad2* and *Smad5*) implicated with the known target *Zeb2* in suppression of epithelial gene expression. To validate these genes as miR-200c targets, we used an optimized version of the PD protocol[39] coupled with qRT-PCR to measure their enrichment for binding to Bi-miR-200c vs Bi-cel-miR-67 as a control. Bi-miR-200c PD in 4TO7 cells significantly enriched 14 of the 18 genes (Figure 13a). The 4 genes that did not confirm (*Skil*, *Bim/Bcl2l11*, *Mapk1* and *Bax*) had lower enrichment ratios (2.3-3.0). As a control, the housekeeping mRNAs *Sdha* and *Tuba1a* were not enriched in the Bi-miR-200c PD.

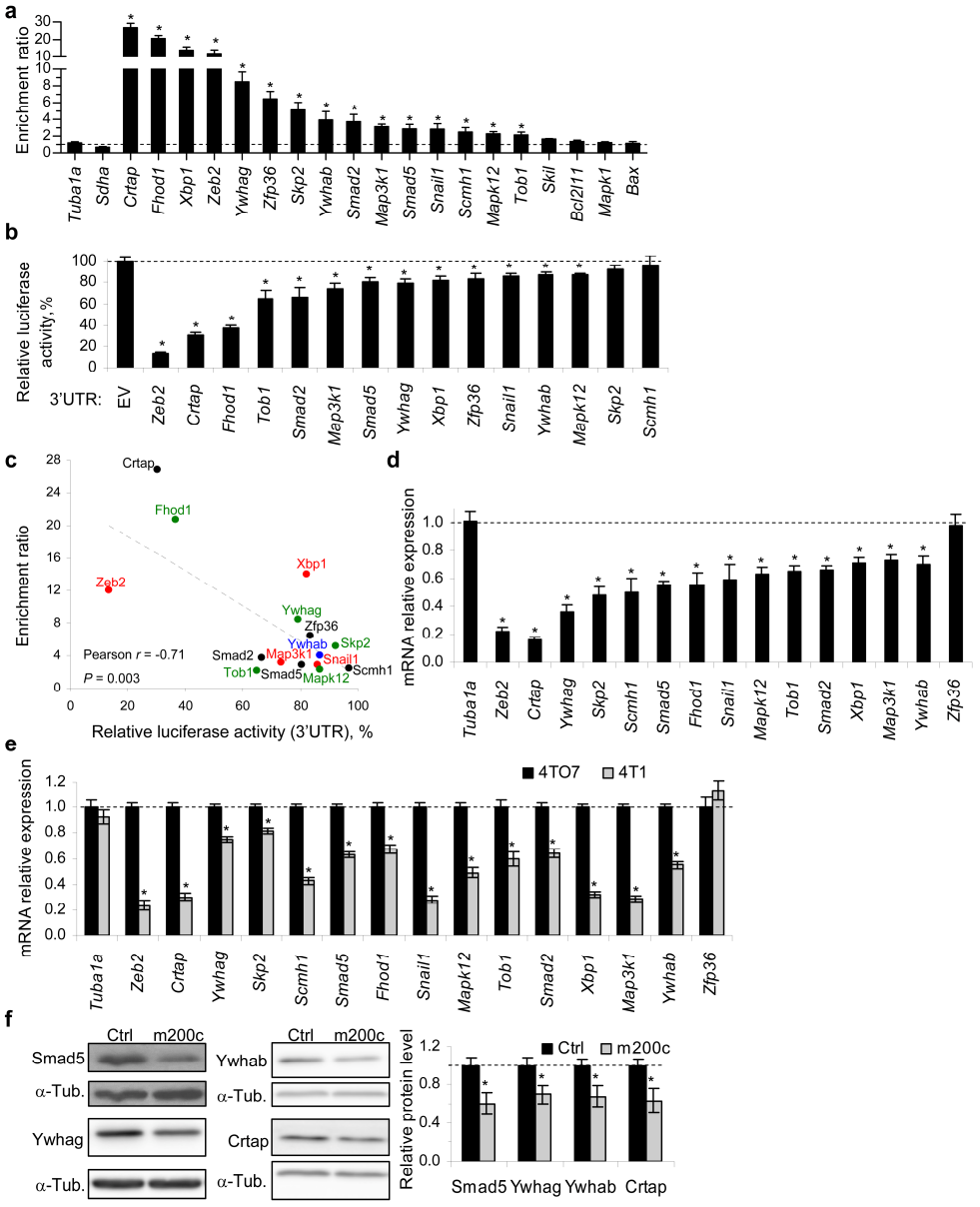


Figure 13. Experimental validation of miR-200c target genes identified by the PD (a) Bi-miR-200c enrichment ratio R in 4TO7 cells 24 h after transfection with Bi-miR-200c or Bi-cel-miR-67 mimics by qRT-PCR normalized to *Gapdh* for 18 genes chosen for validation. R = {miR-200c PD (gene mRNA/*Gapdh* mRNA)}/{Cel-miR-67 PD (gene mRNA/*Gapdh* mRNA)}. *Tuba1a* and *Sdha* are negative controls and *Zeb2* is a positive control. (b) miR-200c regulation of the 3'UTR of the 14 putative target genes confirmed to be pulled down by Bi-miR-200c. 4T1E cells, co-transfected with the dual luciferase reporter psiCHECK-2 bearing the full 3'UTR of each gene in the 3' UTR of the Firefly reporter and with control or miR-200c mimics, were analyzed by luciferase assay 24 h later. Firefly luciferase activity was

normalized to *Renilla* activity. Results obtained for miR-200c transfection were normalized to control miRNA transfection. (c) qRT-PCR PD enrichment ratio (a) correlates inversely with 3'UTR luciferase assay (b). (d) Effect of miR-200c over-expression on mRNA levels of the 14 confirmed pulled down genes, assessed by qRT-PCR normalized to *Gapdh* 48 h after transfection of 4TO7 cells with miR-200c vs control miRNA mimics. (e) qRT-PCR comparison of miR-200 target gene mRNA expression, relative to *Gapdh*, in 4TO7 and 4T1 cells. (f) Immunoblot of total cell lysates extracted 48 h after transfecting 4TO7 cells with control (Ctrl) or miR-200c (m200c) mimics. α-Tubulin is a loading control. The bar graph shows the relative ratio of the signal (normalized to α-Tubulin) in replicate experiments by densitometry, relative to cells transfected with control mimic. Bar graphs represent mean ± SD of three independent experiments. *, $P<0.05$.

To determine whether the confirmed pulled-down genes are direct miR-200 targets, we cloned their full 3'UTRs into a dual-luciferase reporter vector and measured changes in luciferase activity in 4T1E cells after co-transfection with control or miR-200c mimic. miR-200c significantly repressed the 3'UTRs of 12 of the 14 genes by ~15-85% (Figure 13b), confirming that their 3'UTRs are directly regulated by miR-200c. The degree of repression of luciferase activity by miR-200c acting on the 3'UTRs strongly correlated with their enrichment ratio in the PD, suggesting that those mRNAs that bound most strongly were the most highly regulated (Figure 13c). We next looked at whether miR-200c over-expression significantly reduced mRNA levels of the 14 genes in 4TO7 cells, which do not express endogenous miR-200, 48 h later. mRNAs for 13 of the 14 genes were significantly reduced by 25-85% (Figure 13d). To confirm this result in another cell line, we looked at whether miR-200c over-expression significantly reduced mRNA levels of the 12 of these genes expressed in NIH/3T3 cells 48 h later. mRNAs for all 11 target genes that confirmed in 4TO7 were significantly reduced by 15-80% (Figure S1b). (The *Zfp36* gene

that didn't confirm in 4TO7 also wasn't suppressed in 3T3 cells.) The confirmed genes included the two genes (*Scmh1* and *Skp2*) whose 3'UTRs were not regulated by miR-200 over-expression. These might contain miR-200c miRNA recognition elements (MREs) in the CDS. However, this possibility was not examined experimentally. We also compared expression of the 14 target genes in 4TO7 and 4T1 cells. 4T1 cells are fully metastatic, express high levels of miR-200b/c/429 and have an epithelial-like morphology. In contrast, 4TO7 cells are poorly metastatic, don't express the miR-200 family and have a mesenchymal-like morphology[4]. Expression of the 13 of the 14 genes that were validated as targets was significantly downregulated in 4T1 cells compared to 4TO7 cells (Figure 13e). For 4 genes (*Crtap*, *Smad5*, *Ywhag* and *Ywhab*), we also examined the effect of miR-200c over-expression on protein levels 48 h later (Figure 13f). All 4 proteins were significantly reduced after miR-200c over-expression.

To confirm that the 12 genes whose 3'UTRs were regulated by miR-200c were bona fide targets, we next wanted to determine whether their 3'UTRs contain miR-200c MREs. We used TargetScan[47], PITA[50] and RNAhybrid[51] algorithms to identify potential MREs in their 3'UTRs (Table S5). For all 12 genes, except *Smad2*, we identified one MRE, all but one of which had exact seed pairing and verified that it was functional by showing that mutation of its seed sequence within the full 3'UTR restored luciferase reporter activity (Figure 14).

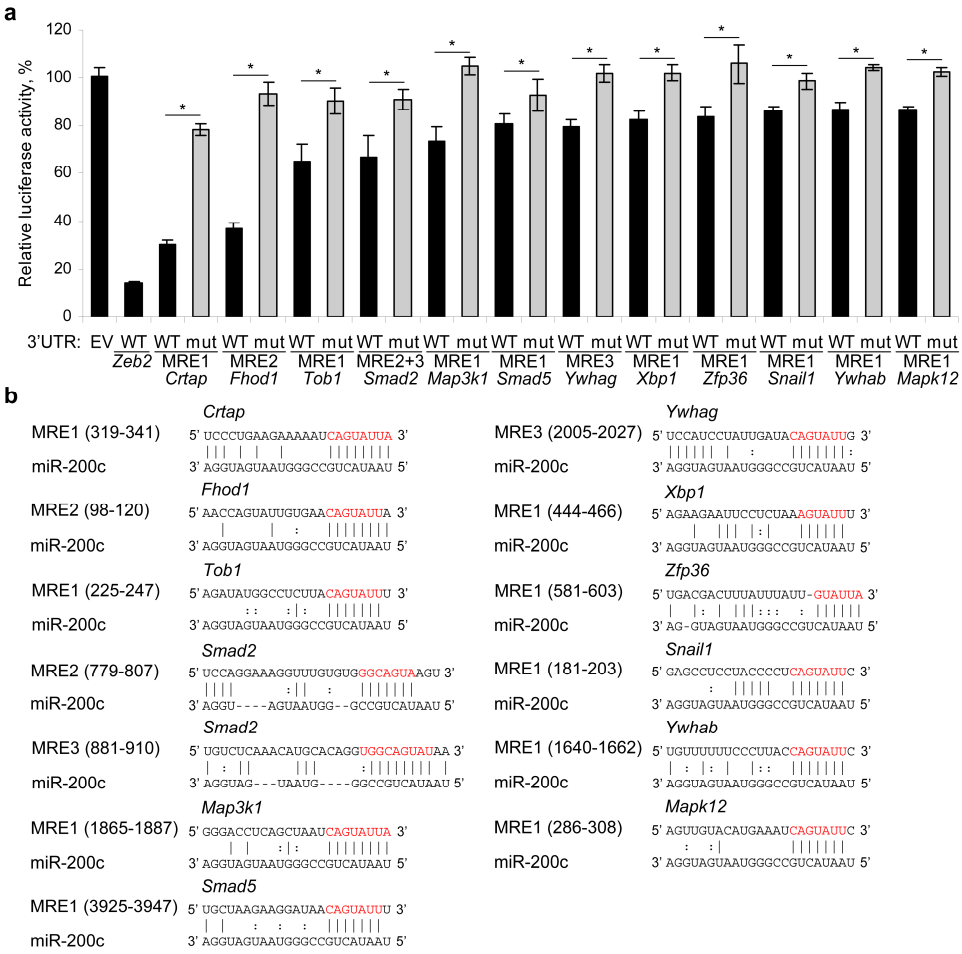


Figure 14. Identification of miR-200c microRNA recognition elements (MREs) in the 3'UTRs of 12 new target genes (a) 4T1E cells, co-transfected with control or miR-200c mimics and with a psiCHECK-2 luciferase reporter plasmid bearing the wild-type full 3'UTR of each gene or a mutated 3'UTR (mut) in which the miR-200c seed region of one or two MREs was mutated were analyzed for luciferase activity 24 h later. Firefly luciferase activity was normalized to *Renilla* activity and results obtained for miR-200c mimic transfection were normalized to control mimic. Data represent mean ± SD of three independent experiments. *, $P<0.05$. (b) Functional miR-200c MREs in each 3'UTR are shown (numbers in parentheses indicate the position of the MRE in each 3'UTR). Point mutations in residues in red were introduced in each full 3'UTR in the mutant plasmid. Perfect matches are indicated by a line; G:U pairs, by a colon. Table S5 shows all the predicted MREs that were tested.

For *Smad2* we identified 2 MREs, neither of which had canonical seed pairing, but both had at least 7 continuous exact matches to miR-200c

overlapping with the seed; mutation of these pairing sequences in both MREs restored luciferase activity, but each individually only partially restored activity (data not shown). The results in Figures 13 and 14 taken together demonstrate that the miR-200c PD is highly specific for identifying miR-200c regulated genes. The low false positive rate in our data set suggests that miR-200c can potentially regulate hundreds of genes.

3.5. miR-200 targets Snail1 and Zeb2 transcription complexes

To confirm that miR-200 enhances epithelial gene expression in 4TO7 cells, we measured the effect of miR-200c over-expression on *Zeb2* and epithelial gene (*Cdh1*, the gene encoding E-cadherin; *Cldn3* and *Cldn7*, claudin genes) mRNAs by qRT-PCR (Figure 15a). As expected, *Zeb2* mRNA significantly declined, while the epithelial gene mRNAs increased 3-10-fold. To determine how much of the effect of miR-200 could be attributed to *Zeb2* knockdown (KD), we compared the effect of miR-200 over-expression and *Zeb2* KD, both of which caused a comparable reduction in *Zeb2* mRNA. *Zeb2* KD only induced the epithelial genes by ~2-fold. Thus miR-200 regulation of other genes likely contributes substantially to promoting epithelial transition.

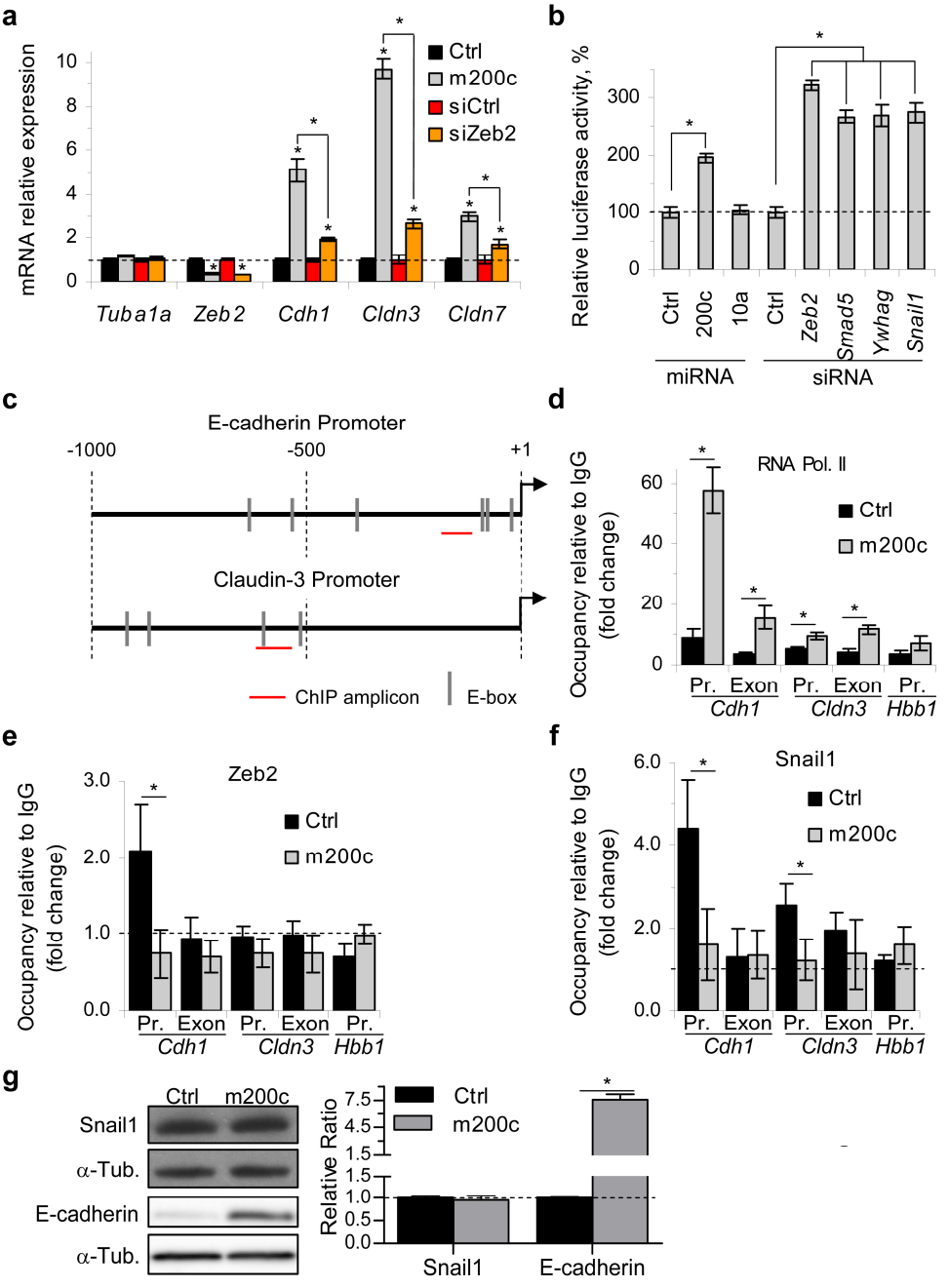


Figure 15. miR-200 induces epithelial gene expression by repressing Zeb2 and Snail1 complexes (a) miR-200c over-expression enhances epithelial gene expression more than *Zeb2* KD. Gene expression in 4TO7 cells was analyzed by qRT-PCR, relative to *Gapdh*, 48 h after transfection with the indicated siRNA or miRNA mimics, relative to control siRNA or miRNA transfection. (b) Effect of miR-200c over-expression and *Zeb2*, *Smad5*, *Ywhag* and *Snail1* KD on the transcriptional activity of a luciferase reporter plasmid bearing the -178 to

+92 fragment of the *Cdh1* promoter. 4TO7 cells were co-transfected with the Firefly luciferase promoter construct and a control *Renilla* reporter vector and 24 h later with the indicated siRNA or miRNA mimics. Relative luciferase activity measured 48 h after the last transfection were normalized to that in cells transfected with the respective control siRNA or miRNA. (c) Schematic of putative Zeb2 and Snail1 binding sites (E-boxes) in the 1000 bp genomic sequence upstream of the transcription start sites (nucleotide +1) of the mouse *Cdh1* (E-cadherin) and *Cldn3* (claudin-3) genes. Sequences amplified by qPCR are indicated in red. (d-f) Binding of RNA polymerase II, Zeb2 or Snail1 to sequences in the *Cdh1* or *Cldn3* promoter (Pr) or exon 24 h after transfecting 4TO7 cells with control (Ctrl) or miR-200c (m200c) mimics was analyzed by ChIP. Graphs show the binding to the specific antibody compared to binding to the same amount of IgG as measured by qPCR. The *Hbb1* (β-globin) promoter region was used as a control. (g) Immunoblot analysis of total protein extracted 48 h after 4TO7 cells were transfected with control (Ctrl) or miR-200c (m200c) mimics. α-Tubulin is a loading control. The bar graph shows the relative ratio of the signal (normalized to α-Tubulin) in replicate experiments by densitometry, relative to cells transfected with control mimic. Bar graphs show mean ± SD of three independent experiments. *, $P<0.05$.

In addition to *Zeb2*, miR-200c pulled down the E-cadherin transcription repressor *Snail1* and the corepressors *Smad2*/*Smad5* and *Ywhag*/*Ywhab* (Figure 12 and Figure 13). Moreover the mRNAs of all of these genes were significantly down-regulated by miR-200 over-expression in 4TO7 cells. The other Smad and 14-3-3 genes were not pulled down with miR-200c and their mRNA levels were not reduced by miR-200c over-expression (Table S2, Figure S2a). To examine whether suppression of these other miR-200c target genes contributes to epithelial gene expression, we co-transfected 4TO7 cells with miR-200c or control miRNA mimics or siRNAs against *Zeb2, Snail1*¸ *Ywhag* or *Smad5* and a pGL3 reporter vector containing the Firefly luciferase gene under the control of the –178 to +92 fragment of the mouse E-cadherin promoter (Figure 15b). Depletion of endogenous *Snail1*, *Smad5* and *Ywhag* individually (verified by qRT-PCR, data not shown) enhanced transcription from the E-cadherin promoter about ~2.5-fold (almost as much as KD of *Zeb2*, which increased promoter

activity ~3-fold) and more than the 2-fold increase in promoter activity caused by miR-200 over-expression.

Zeb2[60] and Snail1[61] are zinc finger transcription factors that bind to a consensus E-box hexanucleotide sequence (CACCTG) in the promoters of epithelial genes to repress their expression. *Cdh1* and *Cldn3* promoters have several E-box sequences (Figure 15c) that render them transcriptional targets of Zeb2 and Snail1. To determine whether miR-200 regulates the binding of endogenous Zeb2 and Snail1 to these promoters and transcription of epithelial genes, we performed Zeb2, Snail1 and RNA pol II chromatin-immunoprecipitation (ChIP) assays one day after 4TO7 cells were transfected with control or miR-200c mimics. miR-200 over-expression significantly enhanced RNA pol II occupancy of the *Cdh1* and *Cldn3* promoter and exonic regions, but not of the *Hbb1* (β-globin) promoter used as a control (Figure 15d), as expected, consistent with enhanced transcription of epithelial genes. In control cells, Zeb2 bound to the *Cdh1* promoter and Snail1 bound to both the *Cdh1* and *Cldn3* promoters, but neither bound above background to exonic regions of these genes (Figures 15e and 15f). miR-200 over-expression completely suppressed Zeb2 and Snail1 binding to these promoters. Thus miR-200c over-expression enhances transcription of epithelial genes, likely by reducing binding of both the Snail1 and Zeb2 transcriptional suppressor complexes.

Zeb2 is a well validated strong target of miR-200c, which we confirmed in this mouse system by showing 80-90% reduction in luciferase reporter activity of the 3'UTR (Figure 13b) and mRNA level (Figure 13d) after miR-200 over-expression. We previously showed that Zeb2 protein becomes undetectable in 4TO7 cells upon miR-200c-over-expression[4], so it is not surprising that Zeb2 binding to epithelial gene promoters became undetectable. Changes in *Snail1* mRNA level and luciferase activity after miR-200c over-expression and Snail1 miR-200c binding were much less than for Zeb2 (Figure 13a-d), suggesting that it is a weaker target. In fact, miR-200c over-expression had no significant effect on Snail1 protein level (Figure 15g), presumably because other mechanisms compensated. Thus reduced binding of Snail1 to the epithelial gene promoters (Figure 15f) was not due to a reduction in Snail1 protein. However, Snail1 binding to E-boxes and its repressor activity requires it to bind to Ywhag or Ywhab, since removing the Snail1 14-3-3 binding domain completely blocks its binding and repressor activity.[42] Ywhag and Ywhab protein levels declined with miR-200c-over-expression (Figure 13f). Reduced 14-3-3 protein levels by miR-200c targeting likely contributed to the absence of detectable Snail1 binding to epithelial gene promoters, although other unknown mechanisms may have contributed.

3.6. miR-200 inhibits the EMT-promoting effects of TGF-β and BMP-2

Genes that mediate TGF-β signaling, which promotes mesenchymal properties, were significantly enriched in the miR-200 target network (Fig. 1d). Both *Smad2 and Smad5*, the receptor-Smads (R-Smad), which we validated as miR-200 targets (Figures 13 and 14 and S2a), were phosphorylated in response to TGF-β treatment of 4TO7 and 4T1 cells, as has been previously reported[62] (Figure S2b). To investigate the effect of miR-200 on TGF-β signaling, we treated 4TO7 and 4T1 cells transfected with miR-200c or control miRNA with TGF-β or control vehicle. In both cell lines, TGF-β significantly reduced *Cdh1* expression as expected and also up-regulated *Zeb2.* Cells over-expressing miR-200 significantly inhibited these TGF-β-mediated effects in both cell lines (Figure 16a). Bone morphogenic proteins (BMPs) belong to the TGF-β superfamily, but mediate their signaling via R-Smads 1, 5 and 8. To study the effect of miR-200 on BMP signaling further, we treated 4TO7 cells transfected with miR-200c or control miRNA with BMP-2. BMP-2 treatment, as expected, significantly decreased the expression of 2 epithelial genes (*Cdh1* and *Cldn3*) and increased phosphorylation of the BMP-activated R-Smads (Figure 16b and c). miR-200 over-expression also significantly inhibited both these effects and also decreased Smad5 protein (Figure 16b and c). miR-200 didn't affect *Smad1* and *Smad8* levels (Figure S2a). Reduced Smad1/5/8 phosphorylation was recapitulated by knocking down *Smad5*

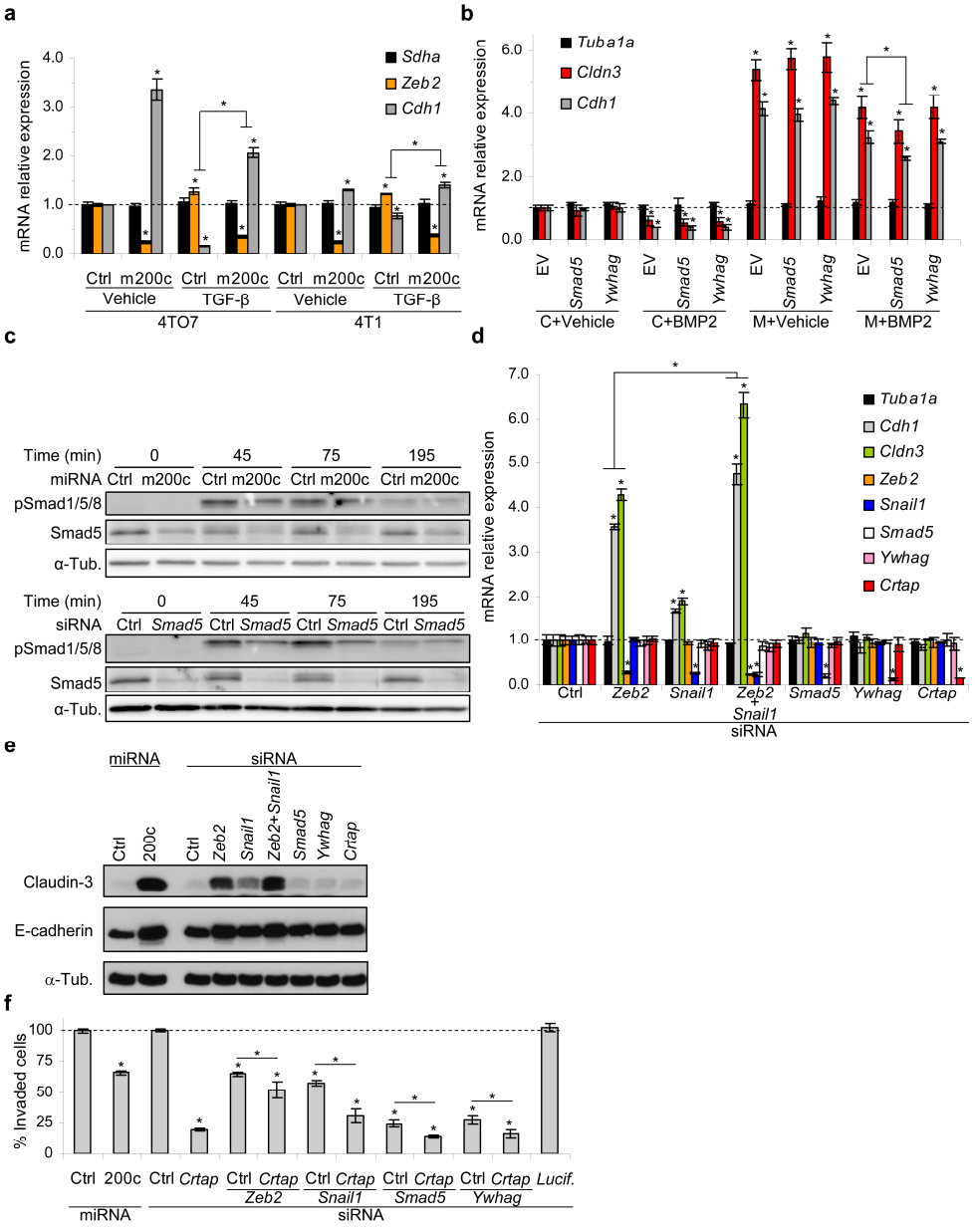


Figure 16. *Snail1*, *Smad5*, *Ywhag* and *Crtap* are physiologically important miR-200c targets

(a) Effect of miR-200c over-expression on TGF-β signaling. qRT-PCR data were normalized to *Gapdh*. 4TO7 and 4T1 cells were transfected with control (Ctrl) or miR-200c (m200c) mimics in the presence of TGF-β or vehicle. RNA was harvested 48 h after transfection. For each cell line, results were normalized to transfection with control miRNA and treatment with vehicle. Asterisks above bars indicate significant differences in relative mRNA expression compared to control miRNA-transfected, untreated samples, while asterisks above brackets indicate significant differences between paired control miRNA and miR-200-transfected

samples. (b) Effect of over-expression of miR-200-resistant *Smad5* and *Ywhag* or empty vector control (EV) on miR-200c-mediated epithelial gene mRNA upregulation in cells treated with vehicle or BMP-2. qRT-PCR data were normalized to *Gapdh.* 4TO7 cells were co-transfected with the indicated pcDNA3.1 expression vectors and control (C) or miR-200c (M) mimics in the presence of BMP-2 or vehicle. RNA was harvested 48 h after transfection. All results were normalized to control conditions in which cells were co-transfected with pcDNA3.1 EV and control miRNA and treated with vehicle. Asterisks above bars indicate significant differences in relative mRNA expression compared to the vehicle-treated sample co-transfected with EV and control mimic, while an asterisk above a bracket indicates significant differences between paired EV and expression plasmid-transfected samples. (c) Effect of miR-200c over-expression (top) or *Smad5* KD (bottom) on BMP signaling. Immunoblot of total protein extracted at indicated times after BMP-2 stimulation of serum-starved 4TO7 cells that were transfected with the indicated miRNA or siRNA 48 h earlier. α-Tubulin is a loading control. (d) Effect of *Zeb2*, *Snail1*, *Smad5*, *Ywhag* and *Crtap* KD on epithelial gene expression. mRNA was analyzed by qRT-PCR relative to *Gapdh* 72 h after transfection of 4TO7 cells with indicated target gene siRNAs. Asterisks above bars indicate significant differences in relative mRNA expression compared to control siRNA-transfected samples, while asterisk above brackets indicates significant difference when *Snail1* KD is combined with *Zeb2* KD. (e) Immunoblot of total protein extracted 72 h after 4TO7 cells were co-transfected with control (Ctrl) or miR-200c (200c) mimics and indicated target gene siRNAs or control (Ctrl) siRNAs. α-Tubulin is a loading control. (f) Invasion through Matrigel of 4TO7 cells co-transfected with miRNA mimics or 50 nM target gene siRNAs and an siRNA against *Crtap* or control (Ctrl) siRNA. In (d-f) 50 nM of each siRNA was used. Ctrl siRNA was added if needed to achieve a final total siRNA concentration of 100 nM. Results were normalized to control siRNA or miRNA transfection. An siRNA targeting Firefly luciferase (Lucif.) was used as a control. Bar graphs show mean ± SEM of three independent experiments. Bar graphs show mean ± SD of three independent experiments, unless otherwise specified. *, $P<0.05$.

(Figure 16c), suggesting that miR-200 represses BMP signaling by suppressing its signaling mediator, *Smad5*. Thus miR-200 inhibits both TGF-β and BMP signaling.

3.7. *Smad5* gene regulation contributes to miR-200 enhancement of epithelial gene expression

The experiments performed so far have involved miR-200c over-expression, which can sometimes exaggerate the role or mechanism of endogenous miRNAs. To investigate the contribution of targeting individual gene targets to miR-200-mediated MET, we first assessed the effect on epithelial gene expression of knocking down some of the epithelial gene transcriptional repressor and corepressor genes in 4TO7 cells. KD of *Zeb2*

or *Snail1* increased E-cadherin and claudin-3 mRNA and protein expression as expected (Figures 16d and 16e), and knocking down both was additive. *Smad5* or *Ywhag* KD did not affect epithelial gene expression, probably because other family members, known to have overlapping functions, may have substituted. The effect of miR-200c over-expression and *Zeb2*, *Snail1*, *Smad5*, *Ywhag* and *Crtap* KD on E-cadherin expression and cell morphology of 4TO7 cells was also analyzed by fluorescence microscopy (Figure S3). Over-expression of miR-200c in 4TO7 cells increased E-cadherin protein and changed 4TO7 morphology from spindle-shaped to cobblestone-forming epithelial-like cells. Although there was no clear microscopic effect of *Snail1* KD, KD of *Zeb2* increased E-cadherin protein expression and epithelial morphology, and knocking down *Zeb2* together with *Snail1* was additive. KD of *Smad5*, *Ywhag* or *Crtap* individually did not affect E-cadherin protein expression, consistent with qRT-PCR and immunoblotting analyses.

To probe the contribution of Ywhag and Smad5 suppression in miR-200c-mediated epithelial gene activation, we cotransfected 4TO7 cells with miR-200c or control miRNA and miR-200c-resistant expression plasmids for these genes. Ywhag and Smad5 protein increased in these cells even in the presence of exogenous miR-200c (Figure S2c). BMP-2, which increased phosphorylation of Smad1/5/8 (Figure 16c), reduced epithelial gene expression and the induction of epithelial genes by miR-200 (Figure

16b). Expression of miR-200-resistant *Smad5* slightly, but significantly, reduced miR-200c-mediated E-cadherin and claudin-3 upregulation in cells treated with the Smad5 activator (BMP-2), but not in cells not treated with BMP-2. Thus, miR-200 suppression of endogenous Smad5 contributes modestly to promoting the MET. However, over-expression of miR-200-resistant *Ywhag* had no significant effect in this assay.

3.8. The miR-200c target *Crtap* promotes cell invasion

Crtap was one of the strongest miR-200c targets, based on its enrichment in the PD (~34-fold) and reduction in luciferase activity and mRNA expression (Figure 13). Not much is known about Crtap. Crtap is an enzyme cofactor that facilitates prolyl-hydroxylation of type I collagen, and its homozygous mutation leads to osteogenesis imperfecta, a severe defect in bone mineralization. Of note, a recent study suggests that prolyl-hydroxylation of type I collagen is required for collagen binding to decorin, a proteoglycan that captures TGF-β in the extracellular matrix. KD of *Crtap* had no effect on epithelial gene expression (Figures 16d and 16e). One of the hallmarks of mesenchymal cells and TGF-β stimulated cells is their motility and ability to invade basement membranes. However, Crtap has no known role in cell motility. To determine whether suppression of *Crtap* or the epithelial gene suppressor targets of miR-200c promote invasion, we knocked them down in 4TO7 cells and measured the effect on cell invasion across a collagen-containing Matrigel-coated membrane (Figure 16f) (KD of

these genes had no effect on cell viability (Figure S2d). KD of *Zeb2* reduced invasivity comparably to transfection of miR-200c. KD of *Smad5*, *Ywhag* or *Snail1* inhibited invasivity even more strongly than either miR-200 over-expression or *Zeb2* KD. However, KD of *Crtap* had the most potent inhibitory effect on invasion. These results suggest that Crtap and other miR-200-regulated gene products have unexpected roles in cell motility/invasion.

3.9. *CRTAP* and miR-200 expression are inversely correlated in NCI-60 tumors and primary human breast cancers

To further study the potential relationship between *Crtap* and miR-200c in cancer, we analyzed the correlation between *CRTAP* and miR-200c expression in the NCI-60 panel of human cancer cell lines. The miR-200 MRE we identified in the mouse gene is evolutionarily conserved in mammals, and human *CRTAP* is a high confidence TargetScan-predicted miR-200c target, which contains a perfect 8-mer seed binding site (Figure S4a). *CRTAP* expression was almost as strongly inversely correlated with miR-200c expression in the NCI-60 panel of human cancer cell lines (Pearson r = -0.44, $P<0.05$) as *ZEB2* expression (r =-0.50, $P<0.05$). The housekeeping gene *SDHA* used as a control was not significantly correlated. We also analyzed data from a study that performed genome-wide mRNA and miRNA profiling in 101 human primary breast cancer samples[63]. *CRTAP* and *ZEB2* expression were also significantly inversely correlated with miR-200c expression in that cohort (Figure S4b). These

findings suggest that *CRTAP* is a physiologically relevant miR-200 target in human cancer.

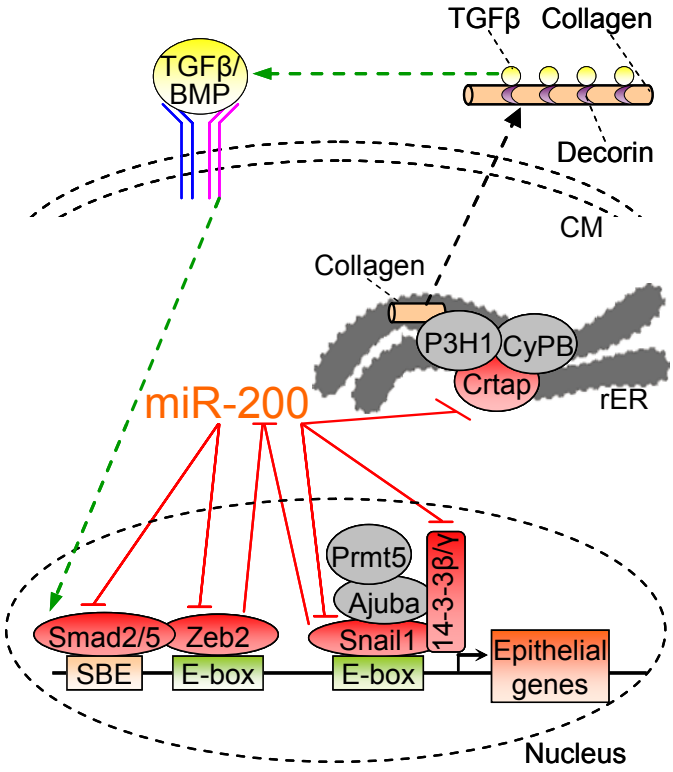


Figure 17. Schematic model of how miR-200c induces epithelial gene expression. miR-200c inhibits the expression of multiple members of the Zeb2 and Snail1 epithelial gene suppressor complexes that are activated by BMP or TGF-β signaling. Both of these complexes also suppress miR-200 transcription in a negative feedback loop[64-65]. Additionally, miR-200 knocks down *Crtap*, which is required for type I collagen prolyl-hydroxylation by prolyl 3-hydroxylase (P3H1) in complex with cyclophilin B (CyPB) in the endoplasmic reticulum. This post-translational modification is required for collagen to bind to decorin, which sequesters TGF-β in the extracellular matrix. Thus, miR-200 KD of *Crtap* may change TGF-β capture by the extracellular matrix and alter TGF-β-stimulated EMT and invasivity. CM, cell membrane; rER, rough endoplasmic reticulum; SBE, SMAD-binding element.

4. Discussion

In this study, we used a biochemical method with high specificity to identify without bias candidate genes that bind to Bi-miR-200c as putative miR-200c target genes. 520 gene mRNAs were enriched at least 2-fold in binding to miR-200c with a $P<0.01$ and 196 were similarly enriched with greater confidence ($P<0.001$). Gene expression of these biochemically identified miR-200 putative target genes was much more significantly downregulated (P=E-6) after miR-200c over-expression than genes predicted to be miR-200c-regulated genes by the best available prediction algorithm (evolutionarily conserved TargetScan targets). The most significantly enriched pathways were EGFR and other cell surface receptor signaling pathways and the MAPK pathways activated secondarily. Importantly, genes involved in TGF-β receptor signaling, which promotes mesenchymal transition and mesenchymal properties, were significantly enriched, and BMP and TGF signaling effects were inhibited by miR-200 over-expression. These target genes, included genes encoding the R-Smad transcription factors, *Smad2* (activated by TGF-β signaling) and *Smad5* (activated by BMP signaling), which we confirmed were miR-200c targets, and *Tgif*, a corepressor of Smad transcription, which we did not experimentally validate. Ectopic expression of miR-200c-resistant *Smad5* slightly but significantly reduced epithelial gene suppression by BMP-2 when miR-200 was ectopically expressed. These results suggest that inhibition of TGFβ/BMP signaling genes by miR-200 contributes to

maintaining epithelial gene expression. miR-200 expression is known to be repressed by BMP-2[66] and TGF-β[67]. Thus our data suggest that miR-200c reinforces the epithelial state not only by inhibiting epithelial gene suppressors, but also by inhibiting TGFβ/BMP signaling.

The miR-200 family is well known to promote epithelial transition through suppression of the epithelial gene transcriptional repressors, *Zeb1/2* and *Snail1/2* (Snail and Slug).[4,18-21,23-24] Here we confirmed that *Zeb2* and *Snail1* were miR-200c target genes (*Zeb1* and *Snail2* were not expressed in these cells). However, although we previously showed that exogenous miR-200c expression strongly suppressed Zeb protein levels in these cells[4], Snail1 protein levels did not change. Snail1 protein was also not downregulated in 4TO7 cells by miR-200c over-expression in a previous study[8]. Although *Snail1* mRNA was consistently down-regulated by miR-200 and an MRE was identified, verifying that it is a bona fide target of miR-200, it was a weak target. Snail1 is a very labile protein with a short half-life (~25 min) regulated by phosphorylation and ubiquitylation[68]. Post-translational modifications of Snail1 protein likely counteracted the weak effect of miR-200 on *Snail1* mRNA. miR-200c orchestrates the expression of epithelial genes not only by suppressing key transcriptional repressors, but also by targeting multiple gene products that act as cofactors in the Zeb and Snail1 repressor complexes, namely Smad2 and Smad5 (with Zeb) and 14-3-3β and 14-3-3γ (with Snail1) (Figure 17).[42] In fact miR-200c-over-expression

had a much stronger effect on epithelial gene expression than *Zeb2* KD, and KD of any of these genes strongly induced the E-cadherin promoter. As a consequence, miR-200 over-expression greatly reduced Zeb2 and Snail1 binding specifically to epithelial gene promoters and enhanced RNA pol II binding to these promoters and transcription of epithelial genes. Snail1 binding to promoters likely depends on the availability of its co-repressors. A previous study showed that KD of the Snail1 co-repressor Ajuba[69] also impairs Snail1 binding without changing Snail1 protein level. KD of other genes in the miR-200c PD, especially *Cbf-A* (*Nfyb*)[70], whose product is a component of the CBF-A/KAP-1/FTS-1 complex that is a master transcriptional regulator of mesenchymal genes, and the NURD complex component *Mta1*[71], which enhances transcription of *Snail1/2* and their recruitment to epithelial gene promoters, may also contribute to promoting the epithelial transition. Thus the ability of miR-200c to foster MET is reinforced by its suppression of multiple components of epithelial gene repressor complexes, its suppression of TGF-β/BMP signaling and likely direct suppression of mesenchymal gene transcriptional activators as well.

We experimentally examined the accuracy of the PD predictions for 18 genes that were not known miR-200c targets when we began this study. 14 of 18 (78%) identified as enriched by the PD and microarray analysis, were found to be enriched by at least 2-fold by more accurate qRT-PCR

analysis. The 4 genes that were not confirmed had a fold-enrichment that was close to our arbitrary cut-off of 2-fold enrichment. We estimated the false positive rate of the PD method used here as 16%, which compares favorably with estimated false positive rates of the best bioinformatics prediction algorithms (which vary between 22 and 31%[72]) or Ago-immunoprecipitation (~40-70%[73]). It is similar to the estimated false positive rate of Ago HITS-CLIP (~13-27%[73]). Our method, which does not use cross-linking, has the advantages that it is simpler and easier to do and requires much fewer cells. The current estimate is similar to an estimate we made for the specificity of a miR-34a PD, where only 1 of 11 (9%) of a random list of putative miR-34a-regulated targets was not verified[34]. That paper required a higher enrichment in the PD (>2.5 fold) than we used here, which could explain the higher specificity. Increasing the cut-off for the microarray data would undoubtedly improve the specificity of the PD method for identifying bona fide target genes, but at the cost of reduced sensitivity. Performing more replicates, and/or changing from microarray analysis of bound mRNAs to RNA sequencing (as we recently described[39]) would likely increase the specificity of the PD. The 3'UTRs of 12 of the 14 confirmed enriched genes (86%) had miR-200-regulated 3'UTRs and we were able to identify 3'-UTR MREs for all of them. Four of these hits were meanwhile independently identified as miR-200-regulated genes by other groups, further supporting our results (*Snail1*[23], *Smad2*[23], *Fhod1*[74] and *Ywhag*[75]). Most of these new 3'UTR MREs (11/13) had exact seed

matches, suggesting that recognition of target genes by miR-200c closely follows canonical seed pairing. The remaining 2 confirmed gene targets may either be false positives or have MREs in the CDS, since recent studies suggest that a large fraction of MREs are in coding regions[73,76-77] and the target gene list was significantly enriched for mRNAs that contain seed pairing sequences in the CDS. The extent of enrichment in the PD correlated well with the extent of miR-200 suppression by luciferase activity (Figure 13c) or mRNA down-regulation after miR-200c over-expression (data not shown). These results suggest that strong enrichment in the PD can be used to identify the set of genes most highly regulated by a miRNA, which may be the more important targets.

The high level of specificity of the PD, coupled with the large number of enriched genes in one cell line that we examined, suggest that miR-200c is capable of regulating hundreds of genes. How many genes are suppressed under physiological conditions is uncertain, since our study relied on miR-200c over-expression where miR-200c levels are not rate-limiting. Moreover, how much the set of regulated genes varies amongst cell types is also unclear for miR-200c or, for that matter, for any individual miRNA. Another open question is how much the target genes for miR-200c are shared with miR-200b and miR-429, which have identical seed sequences or with miR-200a and miR-141, which, although they are considered family

members, actually have seed sequences that differ by one nucleotide. The Bi-miRNA PD method could be used to address these questions.

Crtap, which was enriched in the Bi-miR- 200c PD by ~30-fold by microarray binding to each of 2 probes (Table S2) and by qRT-PCR (Figure 13a) and whose mRNA and protein both were substantially reduced by miR-200c over-expression (by ~6-fold and ~2-fold, respectively; Figure 13d,f), was found to play an unanticipated role in cell invasivity through Matrigel. *CRTAP* expression in human tumor cell lines and a human breast cancer cohort was inversely correlated with miR-200c expression, suggesting that *CRTAP* is also an important miR-200 target in humans. Crtap, which is hypomorphic in some patients with osteogenesis imperfecta (OI), is an endoplasmic reticulum cofactor for prolyl 3-hydroxylation of collagen. Tumor cell secretion of collagen with reduced prolyl 3-hydroxylation may promote tumor invasion. Loss of *Crtap* may reduce collagen binding of small leucine-rich proteoglycans, such as decorin[78], that sequester mature TGF-β within the extracellular matrix. In fact, many of the symptoms of OI in *Crtap*$^{-/-}$ mice can be ameliorated by injection of anti-TGF-β. Changes in TGF-β capture by the extracellular matrix could alter TGF-β-stimulated EMT and invasivity. Another possibility is that modifications of prolines on other unknown secreted substrate(s) are responsible for enhanced invasivity/metastasis of miR-200 over-expressing cells. Further study of the role of Crtap in cancer and metastasis is warranted. Another

highly enriched gene in the PD, which was previously identified as a miR-200c target gene[74], which we showed is a strongly regulated miR-200c target gene, was *Fhod1.* Fhod1 is a formin that bundles and stabilizes actin filaments in lamellipodia and filopodia at the leading edge of moving cells and has been implicated in regulating cytokinesis during mitosis in an E-cadherin dependent pathway. Fhod1 is upregulated in cancer cells and during EMT and has been linked to enhanced invasion and metastasis in some models. In this study we did not look at the functional consequences of *Fhod1* suppression on cancer properties. Because miR-200 expression enhances metastasis in some cancer models[4,7-10] and reduces it in others[11-13], it will be worthwhile to study more the effect of miR-200 on *Crtap* and *Fhod1* in cancer cell migration and metastasis.

5. Note added in proof

After the submission of this manuscript, a report[79] was published, which used Ago-HITS-CLIP to identify candidate miR-200a/b targets in MDA-MB-231 human breast cancer cells. They found that miR-200a/b controls actin cytoskeleton dynamics by regulating gene products involved in Rho-ROCK signaling, invadopodia formation, matrix metalloproteinase activity and focal adhesions. They also identified cell adhesion and TGF-β signaling as among the most enriched pathways for miR-200a/b targets in human cells in agreement with our findings for miR-200c targets in mouse. Integrin-mediated cell adhesion (Figure 1d) and focal adhesion (adjusted *P*=2.1E-

02, data not shown) networks, which share 29 and 48 genes, respectively, with the actin cytoskeleton regulation network were amongst the most-enriched pathways we identify here for miR-200c targets in mouse. About 25% of the target genes we identified for miR-200c in mice were also found by Bracken et al. for miR-200a/b in humans. Of note, 50% of the 12 genes we validated as direct miR-200c targets in mouse by luciferase assay were also identified by Bracken et al. to be miR-200a/b targets (*Crtap*, *Map3k1*, *Smad5*, *Ywhag*, *Zfp36* and *Ywhab*), further supporting our findings. However, we did not identify control of actin cytoskeleton dynamics on its own as a significantly enriched process in miR-200c-regulated mouse cells. This difference in conclusions between our study, which studied a mouse TNBC that has both mesenchymal and epithelial properties, and theirs, which examined a mesenchymal human TNBC cell line, could reflect differences in technique, differences between the relevant miR-200 family members (miR-200c vs miR-200a/b), differences in mouse and human targets, or differences in gene expression between these cell lines. In fact, about half of the mouse miR-200c target genes we found that were not identified as targets of miR-200a/b in human cells do not have TargetScan-predicted miR-200 binding sites that are conserved between mice and humans. Additionally, the Bracken et al. study showed that epithelial human TNBC cell lines, which are not motile, do not express many of the actin motility-related targets identified in the mesenchymal cell line they studied. These differences suggest that some of the target genes and functions of a

miRNA may be shared in different cellular contexts, while others may be context-specific.

6. Conflict of Interest

The authors have no competing financial interests in relation to this work.

7. Acknowledgements

Dr. Fabio Petrocca (Lieberman laboratory) obtained the raw data used to generate Figure 11a.

Bioinformatic analyses in Figures 11a,c,d, S1a and S4b were performed by Dr. Gabriel Altschuler (Department of Biostatistics, Harvard School of Public Health).

Figures 11b, 16e and S3 were obtained by Dr. Marshall P. Thomas (Lieberman laboratory).

Dr. Shen Mynn Tan (Lieberman laboratory) assisted in performing the pull-down experiment in Figure 13a.

Invasion assays in Figure 16f were performed with the assistance of Dr. Minh T. N. Le (Lieberman laboratory).

We also thank Alex Amiet and Devin Leake (Dharmacon) for providing the biotinylated miRNAs, Linfeng Huang (Lieberman laboratory) for assistance with ChIP assays and Francisco Navarro (Lieberman laboratory) for helpful suggestions.

8. Supplemental figures

a

	Adj*P*	nGenes (nPathway)
EGFR1 Signaling Pathway	7.6E-03	7 (167)
Cholesterol Biosynthesis	7.8E-03	2 (12)
Sphingolipid Metabolism	9.1E-03	3 (36)
MAPK signaling pathways	1.1E-02	6 (162)
Diurnally regulated genes with circadian orthologs	1.4E-02	3 (46)
Signaling of HGF Receptor	4.8E-02	2 (31)

b

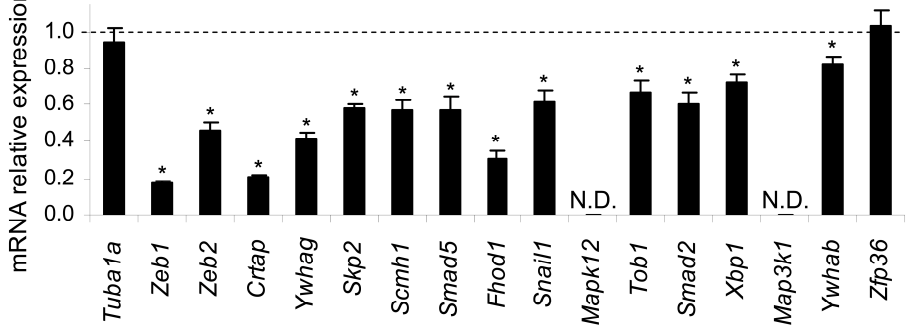


Figure S1. (a) Canonical pathways (Wikipathways) most enriched (adjusted P (AdjP)<0.05) in the Bi-miR-200c PD (R>2, P<0.001). The PD genes in each enriched canonical pathway are listed in Table S4. (b) Effect of miR-200c over-expression on target gene mRNA levels in NIH/3T3 cells. Relative mRNA level, normalized to Gapdh, was assessed 48 h after transfection of miR-200c vs control miRNA mimics. N.D., not detected (transcript not expressed). Shown are mean ± SD of three independent experiments. *, P<0.05.

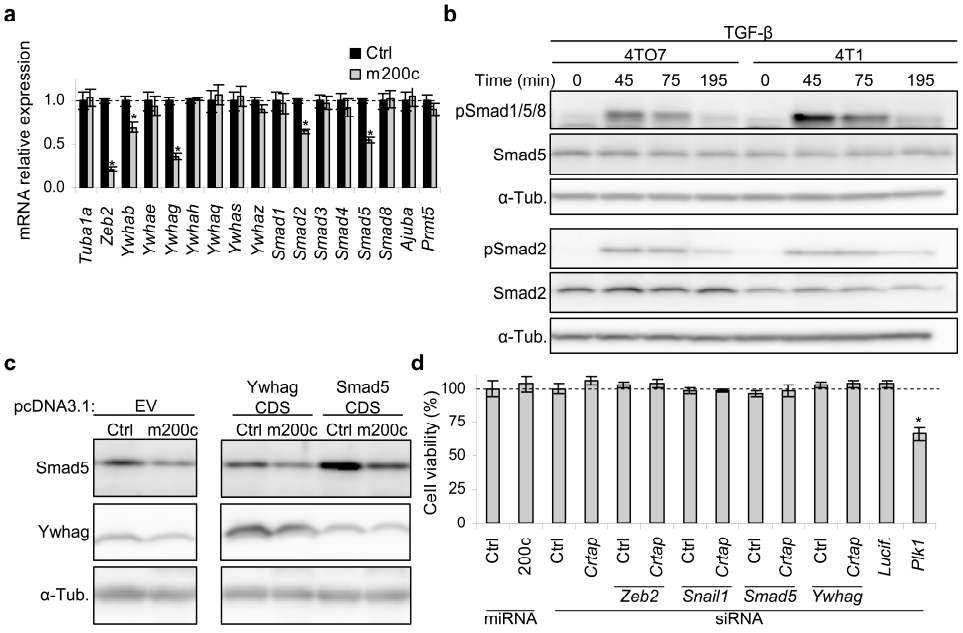


Figure S2. (a) Effect of miR-200c on mRNA expression of 14-3-3 and Smad family genes, Ajuba and Prmt5. qRT-PCR, performed 48 h after 4TO7 cells were transfected with control (Ctrl) or miR-200c (m200c), was normalized to Gapdh. (b) Immunoblot of total protein extracted at indicated times after adding TGF-β to serum-starved 4TO7 or 4T1 cells. α-Tubulin is a loading control. (c) Immunoblot of total cell lysates extracted 48 h after co-transfecting 4TO7 cells with the indicated over-expression vector or empty vector (EV) and control (Ctrl) or miR-200c (m200c) mimics. α-Tubulin is a loading control. (d) 4TO7 cell viability, assessed by CellTiter-Glo 72 h after co-transfection of miRNA mimics or indicated target gene siRNAs and an siRNA against Crtap or control (Ctrl) siRNA. Bar graphs represent mean ± SD of three independent experiments. *, P<0.05.

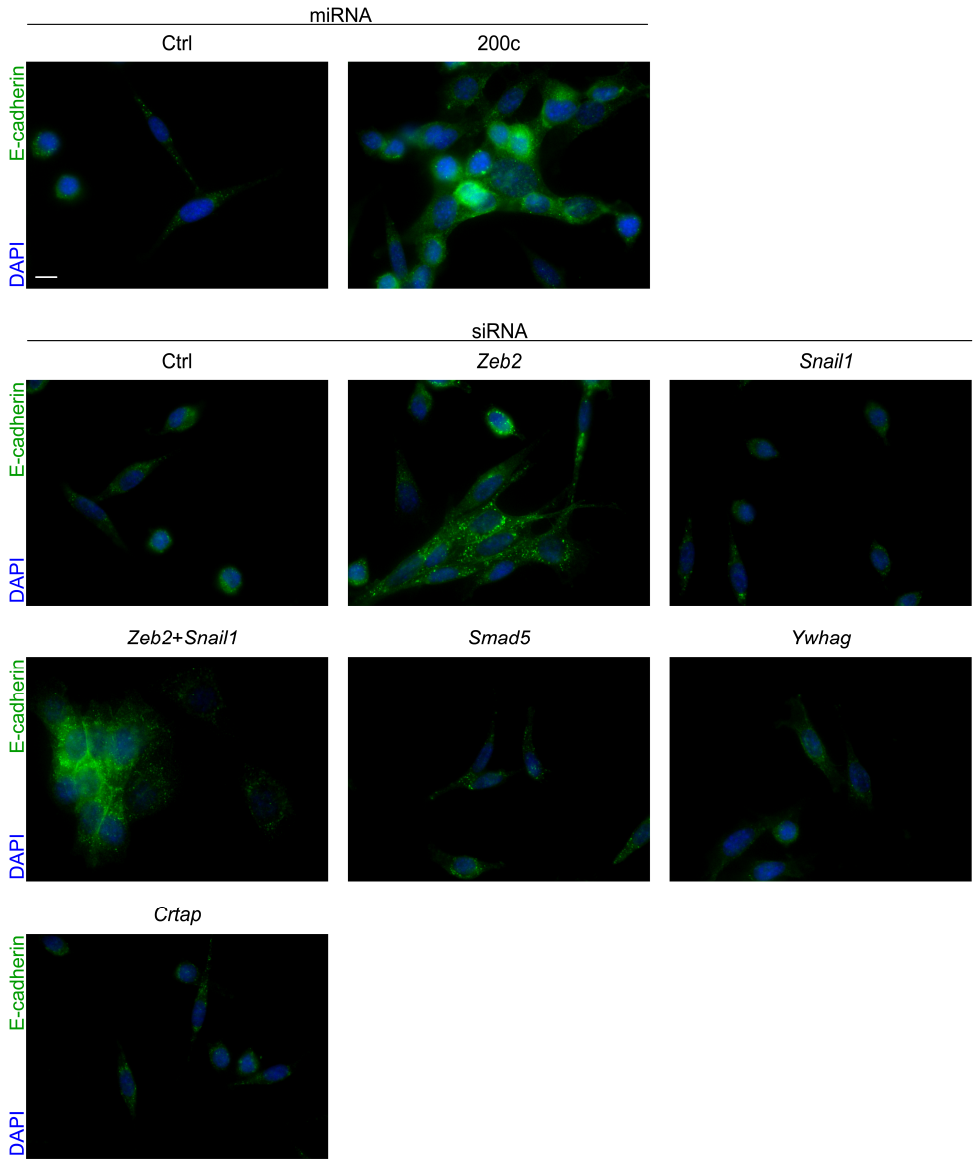


Figure S3. Effect of miR-200c over-expression and Zeb2, Snail1, Smad5, Ywhag or Crtap KD on 4TO7 cell morphology and E-cadherin staining (green). DAPI (blue) was used to stain nuclei. Scale bar,10 μm.

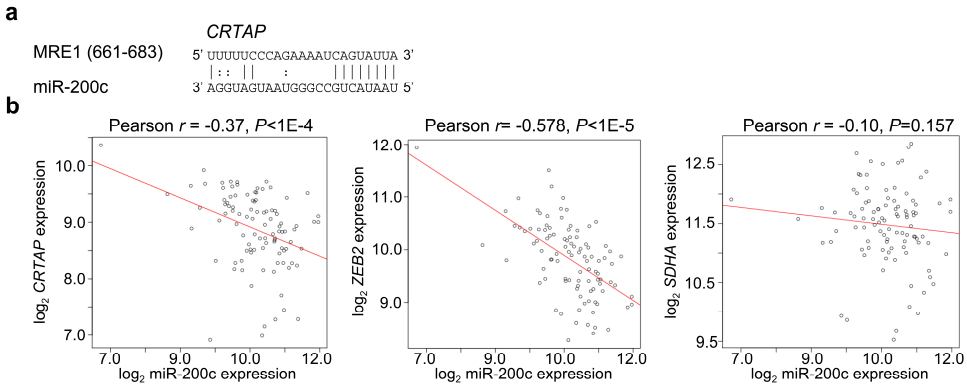


Figure S4. (a) TargetScan-predicted miR-200c MRE in human CRTAP 3'UTR (numbers in parentheses indicate the position of the MRE in the 3'UTR). Perfect matches are indicated by a line; G:U pairs, by a colon. (b) Correlation of miR-200c and CRTAP, ZEB2 and SDHA expression in a cohort of human primary breast tumors.

10. Supplemental tables

Table S1. Published miR-200 target genes and their overlap with the PD gene list. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown. N.D., not detected (transcript not expressed by Bi-cel-miR-67-transfected 4TO7 cells). N.P., no probe (microarray had no probe for transcript).

	Gene	Species	miR-200 family member	PD enrichment-ratio	Ref.
1	ACVR2B	H	200c, 141	1.5	80
2	AP-2A (TCFAP2A)	H	200c, 200b, 429	**2.3**	81
3	ARHGAP19	H	200c	N.D.	82
4	BCL-2	H H	200b, 200c, 429 429	1.0	33 83
5	BMI1	H H H M	200b 200c 141 200c	N.D.	12,26 84-86 87 88
6	BRD3	H	141	N.P.	89
7	BRD7	H	200c	N.P.	90
8	CAGE	H	200b	N.D.	91
9	CD47	H	141	1.9	92
10	CDC25B	H	141	1.1	93
11	CDH11	H M	200c 141	N.D.	94 95
12	CDK2	H	200b	1.7	96
13	CDK6	H	200a	N.D.	97
14	CDKN1B (p27/kip1)	H	200b	N.D.	98
15	CFL2	M H	200c 200b	**24.5**	8 96
16	c-Maf	M	200c	N.D.	99
17	c-MYB	H	200b, 200c, 429	N.D.	100
18	C-MYC	H	429	1.5	101
19	CREB1	H	200b	N.D.	102
20	CTNNB1 (β-CATENIN)	R H H	200a 200a 200a	1.4	103-105 106 107
21	CUL3	H	141	0.95	92
22	CXCL1	H	200a, 200b	**2.1**	108
23	CXCL12β	H, M	141	N.D.	109
24	CYCLIN E2	H	200a, 200b, 200c	N.D.	110
25	DLC-1	H	141	N.D.	111
26	Dlx5	M	141, 200a	N.D.	66
27	DNMT3A	H	200b, 200c	N.D.	112
28	DNMT3B	H	200b, 200c	1.0	112
29	E2f3	M	200c	1.1	113
30	eIF4E	H	141	N.D.	114

Table S1 (cont.). Published miR-200 target genes and their overlap with the PD gene list. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown. N.D., not detected (transcript not expressed by Bi-cel-miR-67-transfected 4TO7 cells). N.P., no probe (microarray had no probe for transcript).

	Gene	Species	miR-200 family member	PD enrichment-ratio	Ref.
31	EPHA2	H	200a	0.53	115
32	EPHRIN-A1 (EFNA1)	H	200c	**3.3**	116
33	ERRFI-1	H	200c	**4.0**	117
34	ETAR (EDNRA)	H	200c	0.93	118
35	ETS-1	M H	200c 200b	1.7	119 22
36	FAP-1 (PTPN13)	H	200c	N.D.	31
37	FBLN5	H	200c	N.D.	120
38	FHOD1	H	200c	**6.9**	117
39	FLK1 (KDR, VEGFR2)	M M H H	200c 200b 200b 200c	N.D.	119 121 122 123
40	FLT1 (VEGFR1)	H H H	200b 200c 200c	**2.5**	122 124 125
41	FN1	H M H	200c 200c/141 200b	N.P.	82 126 127
42	FOG2 (ZFPM2)	H	200a, 200b, 200c 141, 429	N.D.	128
43	FTH1 (Ferritin)	H	200b	1.2	129
44	GATA2	H	200b	**2.6**	121
45	GATA4	H H	200b 200c	N.D.	30 130
46	GNA13	H	200a 141	**2.2**	131
47	Grb2	M	200a	N.D.	98,132
48	HDAC4	H	200a	0.65	133
49	HDGF	H	141	1.7	134
50	IKKB (Ikbkb)	H	200c	0.93	124
51	IL-8 (Cxcl15)	H	200a, 200b	N.D.	108
52	JAGGED1 (Jag1)	H	200c, 141	N.D.	135-137
53	KEAP1	H H M	141 200a 200c/141	0.92	138 139 126

Table S1 (cont.). Published miR-200 target genes and their overlap with the PD gene list. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown. N.D., not detected (transcript not expressed by Bi-cel-miR-67-transfected 4TO7 cells). N.P., no probe (microarray had no probe for transcript).

	Gene	Species	miR-200 family member	PD enrichment-ratio	Ref.
54	KINDLIN-2 (FERMT2)	H	200b	1.9	96
55	KLF4	H	200b	1.2	26
56	KLF9	H	200c	1.0	124
57	KRAS	H	200c	0.76	140
58	LEPR	H	200c	N.D.	82
59	LPAR1	H	200c	N.P.	118
60	Lrp1	M	200c	1.4	8
61	MAML2	H	200c	N.D.	136
62	MAML3	H	200c	N.D.	136
63	MAP4K4	H	141	0.71	141
64	MAPK14 (P38A)	H	200a, 141	0.80	142
65	MARCKS	H	200c	**2.2**	143
66	MKK4	H	141	0.79	144
67	MSN	H H	200c 200b	**9.3**	82 145
68	MUC16	H	200c	N.P.	146
69	MUC4	H	200c	N.D.	146
70	NCAM1	H	200c	N.D.	147
71	Noggin (Nog)	M	200c	N.D.	148
72	NOTCH1	H H	200b, 200c 200b	0.77	149 150
73	NOXA	H	200c	N.D.	151
74	Nrp1	M	200a	N.D.	95
75	NTF3	H	200c	N.D.	152
76	NTRK2 (TRKB)	H	200c	N.D.	82
77	ONECUT2	H	429	N.D.	153
78	OREBP	H	200b	1.09	154
79	OXR1	H	200b	1.6	155
80	P53	H	200a	0.86	156
81	PAF	H	200b	0.94	96
82	PIN1	H	200b	N.D.	157
83	Pkd1 (polycystic kidney disease 1)	M	200a, 200b	**7.1**	158
84	PLCG1	H	200b,200c,429	N.D.	28
85	PPARA	H	141	N.D.	159
86	PPM1F	H	200c	**5.2**	74
87	PRKAR2B	H	200b	N.D.	160

Table S1 (cont.). Published miR-200 target genes and their overlap with the PD gene list. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown. N.D., not detected (transcript not expressed by Bi-cel-miR-67-transfected 4TO7 cells). N.P., no probe (microarray had no probe for transcript).

	Gene	Species	miR-200 family member	PD enrichment-ratio	Ref.
88	PSMD1	H	200b	0.96	161
89	PSMD1	H	200b	0.96	161
90	PTEN	H	141	N.D.	89
		H	200a		162
		H	200c		163
		M	200a		164
91	RAB18	H	429	**4.4**	165
		H	200b		13,166
92	RAB21	H	200b	**3.1**	13,166
93	RAB23	H	200b	N.D.	13,166
94	RAB3B	H	200b	N.D.	13,166
95	RBBP4	H	429	N.P.	167
96	RHOA	H	200c	N.D.	118
97	RND3 (RhoE)	H	200b	**4.1**	29
		H	200b		98
		H	200c		168
98	ROCK2	H	200b, 200c	N.D.	169
99	SEC23A	H	200c	**5.7**	94,170
		M	200c		8
100	Serca2b (Atp2a2)	M	200b	1.3	171
101	SHP (NR0B2)	H	141	N.D.	172
102	SIRT1	H	200a	1.0	173
103	Slc25a3	M	141	1.1	174
104	Slug (Snai2)	M	200b	N.P.	24
105	SMAD2	H	141, 200c	**5.5**	175
		M	200b		23
		H	200b		176
106	Snail (Snai1)	M	200b	**2.3**	23
107	SOX17	H	141, 200a	N.D.	137
		H	200a		177
108	SOX2	H	429	N.D.	10
		M	200c		113
109	SP1	H	200b, 200c	N.D.	112
		H	429		83
110	Srf	R	200b	0.98	178
111	STAT5B	H	200a	N.D.	179
112	SUZ12	H	200b	0.86	26,180
		H	200b, 200c		169
113	TBK1	H	200c	1.1	181

Table S1 (cont.). Published miR-200 target genes and their overlap with the PD gene list. Only genes experimentally validated to be directly targeted by miR-200 were considered. Species (Human (H), Mouse (M) or Rat (R)) in which gene was experimentally validated and miR-200 family members used for validation are shown. N.D., not detected (transcript not expressed by Bi-cel-miR-67-transfected 4TO7 cells). N.P., no probe (microarray had no probe for transcript).

	Gene	Species	miR-200 family member	PD enrichment-ratio	Ref.
114	TGF-B2	H R M R	141 200a 200c/141 200a	N.D.	18 182 126 104
115	TGFBR1	H	141, 200a	0.99	175
116	THBS1	H	200a	N.D.	183
117	TIAM1	H	141	0.75	184
118	TIMP2	H	200b	0.86	185
119	TIMP2	H	200c	0.86	120
120	TM4SF1	H	141	N.D.	186
121	TUBB3	H H	200c 200c	N.D.	187 188
122	UBAP1	H	141	1.1	89
123	UBASH3B	H	200a	N.D.	189
124	UBQLN1	H	200b	1.3	161
125	UBQLN1	H	200b	1.3	161
126	VEGF-A	H, M H, R H H H	200b 200b 200b 200c 200c	**3.9**	190 191 122 124 120
127	WAVE3 (WASF3)	H	200b	N.D.	192
128	Wnt1	M	200b	N.D.	193
129	XIAP	H	200b, 200c, 429	N.D.	33
130	YAP1	H H	200a 141	0.40	194 195-196
131	YWHAG (14-3-3γ)	H	141	**2.4**	197
132	ZEB1	H M	200a, 200b 200b, 200c	N.D.	21 4
133	ZEB2 (SIP1)	H M	200a, 200b 200b, 200c	**3.5**	21 4
134	ZMPSTE24	H	141	1.1	198
135	ZNF217	H	200c	N.D.	199

Table S2. Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
620593	56693	Crtap	33.22289891	1.02E-05
7050056	56693	Crtap	33.68810046	1.02E-05
4200731	57437	Golga7	48.02714017	1.02E-05
6110646	102920	Cenpi	18.25184738	2.31E-05
780242	26987	Eif4e2	19.94589604	2.31E-05
20136	12632	Cfl2	24.51199083	3.77E-05
1090398	15463	Agfg1	14.11147309	3.77E-05
2850014	50996	Pdcd7	12.87986176	3.77E-05
4150747	235293	Sc5d	14.95113518	3.77E-05
6350687	30953	Schip1	13.02135145	3.77E-05
4280053	57743	Sec61a2	12.68069478	3.77E-05
2000041	104725	1110002B05Rik	14.35794069	3.79E-05
7570577	234407	Glt25d1	11.58566241	3.92E-05
5870435	244373	Erlin2	9.520712401	5.75E-05
5360577	98404	AI597479	9.783340291	6.69E-05
430332	56431	Dstn	8.876270519	6.74E-05
6520403	74325	Cltb	11.84292104	6.79E-05
2100537	217615	Ctage5	9.362829844	6.79E-05
3800600	74325	Cltb	12.81466702	6.79E-05
6110438	69786	Tprkb	9.186277535	6.79E-05
130768	116972	2310047D13Rik	7.679200315	7.24E-05
7320162	116972	Fam57a	7.602296549	7.24E-05
3180114	19347	Dennd5a	7.595606923	7.24E-05
5690181	58233	Dnaja4	7.843275237	7.24E-05
5670082	67468	Mmd	9.672890863	7.24E-05
2970554	18201	Nsmaf	8.062782705	7.24E-05
4880035	78600	Pde6h	8.753688404	7.24E-05
3840561	18701	Pigf	8.263484844	7.24E-05
520121	19347	Dennd5a	8.093986252	7.24E-05
6520424	19349	Rab7	7.674612253	7.24E-05
5290121	18983	Cnot7	9.523114355	7.54E-05
3780463	15463	Agfg1	10.34231315	7.54E-05
1010528	13196	Asap1	7.450264451	7.54E-05
1400386	17122	Mxd4	8.15693624	7.54E-05
3190735	59021	Rab2a	7.246960893	7.54E-05
870167	19349	Rab7	10.68476717	7.54E-05
1580653	22232	Slc35a2	9.739732145	7.54E-05
6110132	27058	Srp9	9.044670818	7.54E-05

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
4260544	108927	Lhfp	21.28056787	7.55E-05
1430176	78653	Bola3	9.360663488	8.16E-05
7200168	56428	Mtch2	7.220782687	8.41E-05
20139	192292	Nrbp1	7.044515953	8.45E-05
1110520	74450	Pank2	7.132803968	8.52E-05
150709	11637	Ak2	7.906066037	8.88E-05
1710242	13601	Ecm1	8.002126269	8.88E-05
2490369	18201	Nsmaf	6.921245127	8.88E-05
2970035	67105	1700034H14Rik	6.542570796	8.89E-05
4890327	107566	Arl2bp	13.10371209	8.89E-05
2260471	68337	Crip2	8.419543226	8.89E-05
580682	15980	Ifngr2	7.012866948	8.89E-05
2060162	217615	Ctage5	8.795716759	8.89E-05
1450369	72085	Osgepl1	6.775492411	8.89E-05
520176	18763	Pkd1	7.12348883	8.89E-05
5050192	26936	Mprip	6.809780009	9.01E-05
6280056	56428	Mtch2	6.781643143	9.01E-05
540603	56878	Rbms1	6.08129812	9.01E-05
7550463	67554	Slc25a30	8.584525383	9.22E-05
3360431	108927	Lhfp	6.475258638	9.32E-05
110181	104771	Jkamp	8.107069662	9.35E-05
2450504	27058	Srp9	6.796239793	0.000104009
7570324	223473	Nipal2	6.796311217	0.00010631
3310270	116914	Slc19a2	5.951480509	0.000106694
5420538	108927	Lhfp	6.833107635	0.000114594
2850133	17698	Msn	9.300612375	0.000114594
360725	66917	Chordc1	5.496606703	0.000134838
4050274	17126	Smad2	5.484380137	0.000137289
4010021	106064	AW549877	6.035423702	0.000151789
4830433	70369	Bag5	5.734568778	0.000151789
1990333	76987	Hdhd2	7.283177642	0.000154043
5720241	18645	Pfn2	5.874332926	0.000154043
3890538	107566	Arl2bp	11.06715274	0.0001545
3310037	20403	Itsn2	5.153094516	0.0001545
730327	118451	Mrps2	5.618507159	0.0001545
4050039	68192	Leprotl1	5.290632358	0.000157021
870519	68606	Ppm1f	5.245202624	0.000157021
3400474	26399	Map2k6	5.030075183	0.000162865
5310427	116870	Mta1	6.065842398	0.000162865

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
3870070	57743	Sec61a2	5.217878032	0.000165304
4830288	74763	Nat15	5.933865377	0.000192329
6270603	67105	1700034H14Rik	5.90274568	0.000192329
6840440	12238	Commd3	6.588954516	0.000213656
670113	76987	Hdhd2	4.945811358	0.000213656
3890300	19349	Rab7	4.724972037	0.000213656
5560615	74205	Acsl3	4.922986204	0.000218511
3390687	227154	Stradb	5.933634328	0.000218511
5080348	80517	Herpud2	6.785977698	0.000218511
290672	18645	Pfn2	5.093305958	0.000218511
3930154	232086	Tmem150a	7.367786459	0.000218511
2680372	234686	Fhod1	6.864070779	0.000222253
1300674	26936	Mprip	5.646587014	0.000224573
6250273	227154	Stradb	6.155144831	0.000224573
1710133	11974	Atp6v0e	5.188267071	0.000224573
1450608	16525	Kcnk1	4.656544525	0.000224573
770288	21366	Slc6a6	4.926992876	0.000224573
2900010	101476	Plekha1	4.635681438	0.000225939
580463	226151	Fam178a	4.887157937	0.000226324
4670598	67105	1700034H14Rik	6.056782057	0.000226761
1090400	76987	Hdhd2	5.449436639	0.000226761
6620132	80517	Herpud2	4.710270859	0.000226761
7160328	53885	Nphp1	4.835917925	0.000228311
4280020	56494	Gosr2	14.50022666	0.000239523
2710343	232679	Zc3hc1	5.695254256	0.000239523
4670156	56494	Gosr2	6.158971069	0.000241355
7210440	56772	Mllt11	5.69237131	0.000241355
3610470	19330	Rab18	4.414877695	0.000241355
4040471	27401	Skp2	7.733005571	0.000241355
1230564	22234	Ugcg	4.605526772	0.000241355
1440424	332397	Nanos1	5.588278162	0.000246145
5900445	20437	Siah1a	4.312101706	0.000246631
2710754	232679	Zc3hc1	4.453973799	0.000246631
770746	71910	Ppapdc1b	5.597658416	0.000249739
2480544	22210	Ube2b	6.25388249	0.000249739
4860181	28084	Vps25	4.806940194	0.000249739
5090377	17169	Mark3	4.314137338	0.000256298
6840682	12929	Crkl	4.36328845	0.000261021
5700168	22319	Vamp3	4.99059457	0.000261021

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
4180168	67978	Tctn2	4.82628489	0.000262906
4670653	66279	Tmem218	4.531162273	0.000264145
3520093	69534	Avpi1	5.368952827	0.000264145
1980072	20334	Sec23a	5.69614504	0.000264145
2970685	67657	Rabl3	4.448564314	0.000265025
610762	108679	Cops8	4.374788558	0.000268846
1430068	104771	Jkamp	5.039415539	0.000270935
4040520	211556	Ap1ar	4.613524792	0.000270935
2030593	15481	Hspa8	12.96477817	0.000270935
4150669	57874	Ptplad1	4.265236964	0.000270935
3850131	56516	Rbms2	5.559895369	0.000270935
1110564	69587	Pcgf3	4.102938612	0.000279127
2650494	56737	Alg2	5.545068857	0.000282327
1660037	28084	Vps25	4.393822659	0.000282327
4850020	27401	Skp2	5.545992937	0.000285351
3870474	56210	Rev1	4.690744437	0.000291163
6290332	231999	Plekha8	4.222556034	0.000294994
3450100	22433	Xbp1	4.386164703	0.000294994
620541	271457	Rab5a	4.319166391	0.000295933
3310170	80517	Herpud2	3.997075649	0.000298019
1570328	68465	Adipor2	4.24179422	0.000305846
4880554	98417	Cnih4	5.590864091	0.000305846
4570368	74155	Errfi1	4.025833967	0.000305846
2140059	225020	Fez2	4.246037152	0.000305846
10500	56878	Rbms1	5.522035397	0.000305846
3290142	59040	Rhot1	3.988721556	0.000305846
1400332	77733	Rnf170	4.081627001	0.000305846
3450243	27401	Skp2	5.355364473	0.000305846
1470324	381356	5930434B04Rik	4.819962775	0.000311648
2970609	194126	Mtmr11	3.882154204	0.000317435
3180195	330189	Tmem120b	4.53091871	0.000317435
6220026	22695	Zfp36	4.387324932	0.000319104
2850482	109006	Ciapin1	3.912738153	0.000326635
4610193	226849	Ppp2r5a	3.939804418	0.000326851
7000307	83435	Plekha3	5.63509596	0.000332467
10543	215210	Tmem120a	5.551418127	0.000332467
5690327	75909	Tmem49	5.183617485	0.000332467
130437	17121	Mxd3	4.083921185	0.000333729
2100129	56494	Gosr2	8.914243931	0.000334633

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
1660154	66229	Rpl7l1	3.87939545	0.000342354
620291	13601	Ecm1	4.935333751	0.000351471
4810162	328977	Zfp532	3.86717906	0.000351471
6760088	15980	Ifngr2	5.250525175	0.000392715
4810347	56205	Ensa	4.862774516	0.000404049
4150682	76073	Pcgf5	4.049897787	0.000408337
7000563	12042	Bcl10	3.709597638	0.000410951
4180577	14057	Sfxn1	4.32842507	0.000418186
4210274	54401	Ywhab	3.743605538	0.000423615
1690753	21672	Prdx2	3.686266125	0.000424186
6840075	108159	Ubxn8	3.679538965	0.00043243
6770553	233280	Nipa1	3.936264925	0.000442292
1260093	13559	E2f5	3.851160608	0.000446941
2070682	104771	Jkamp	7.285768481	0.000449528
5860148	66701	Spryd4	3.609761997	0.000458427
290168	20775	Sqle	11.4182398	0.000458427
2320452	225215	Rsl24d1	3.787064491	0.000464861
1660170	171095	Il17rc	3.56869606	0.000464861
3450470	56442	Serinc1	5.276576081	0.000467597
4250131	22057	Tob1	4.790663807	0.000467597
2140687	54401	Ywhab	4.742037304	0.000467597
1440142	233103	4931406P16Rik	3.728686199	0.00047434
4610468	266690	Cyb5r4	3.61526519	0.00047434
130138	72536	Tagap	4.471504094	0.00047434
990537	67878	Tmem33	3.558054623	0.00047434
2680014	54473	Tollip	5.988522405	0.00047434
4050370	26936	Mprip	4.924997258	0.000486011
430717	219144	Arl11	4.861572116	0.000486011
5890504	26431	Git2	3.654498245	0.000488232
5860358	22339	Vegfa	3.909604634	0.000491337
110747	22319	Vamp3	4.62948142	0.000510418
2850440	80517	Herpud2	4.317272039	0.000523844
5860064	232679	Zc3hc1	8.255181587	0.000523844
4760075	67978	Tctn2	3.649521358	0.000529011
780187	11637	Ak2	3.510294828	0.00053505
6130390	76890	Memo1	3.569274376	0.000540019
2760750	381269	Mreg	3.411259612	0.000540019
5910594	66701	Spryd4	4.276082204	0.000540019
2510259	232679	Zc3hc1	3.781014022	0.000540019

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
2120243	67861	Akr1b10	3.689629085	0.000544578
1090577	56428	Mtch2	4.923094227	0.000544578
4070400	74194	Rnd3	4.104603098	0.000544578
3180152	217430	Pqlc3	3.685083005	0.00055343
7210082	228859	Fitm2	4.353843282	0.000555726
2600201	57895	Ccdc126	3.500110608	0.000557855
3180703	106840	Unc119b	5.226992742	0.000557855
1470470	378702	Serf2	3.908527756	0.000558551
360091	69674	Mif4gd	3.91869104	0.000579511
6650215	140546	Eri3	3.455278935	0.000589655
4670370	194126	Mtmr11	6.511425766	0.000611935
3870039	171095	Il17rc	3.417615968	0.000629896
5960594	231506	Lin54	3.481921163	0.000629896
940164	208638	Slc25a38	3.419878238	0.000629896
2370414	104416	Bap1	3.309695075	0.000634044
6660725	18457	Pldn	3.348305564	0.000649679
7610731	69786	Tprkb	3.838240894	0.00065732
5340246	22121	Rpl13a	3.451909692	0.000657678
6330309	108705	Pttg1ip	3.654712814	0.000664888
2370754	217430	Pqlc3	3.241300633	0.000671263
7040133	69534	Avpi1	3.679632122	0.000672833
2140743	545085	Wdr70	3.926387921	0.000682082
160672	328977	Zfp532	3.65878069	0.000682082
4730066	30953	Schip1	7.376623524	0.000737209
4200100	19252	Dusp1	3.491153759	0.000743182
2940209	22193	Ube2e3	3.59990988	0.000743182
7200463	71448	Tmem80	3.371656009	0.000745628
1660075	225215	Rsl24d1	5.835571293	0.000751007
3460026	100047167	LOC100047167	3.659034738	0.000752904
4780671	21665	Tdg	4.065891604	0.000752904
1450364	20471	Six1	4.253343773	0.000753716
1780132	16210	Impact	3.166821688	0.000761857
7570008	27965	Spg21	3.157345704	0.000761857
3170292	215449	Rap1b	3.199829759	0.000791057
2570008	117198	Ivns1abp	3.177879523	0.000792761
2340039	218630	Ccno	3.360540084	0.000793088
7400725	17313	Mgp	4.246552872	0.000796369
2140678	232086	Tmem150a	6.247263126	0.000797362
6400403	192656	Ripk2	4.250945336	0.000811618

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
5960332	71983	Tmco6	4.097802402	0.000814813
7330241	17346	Mknk1	4.67039567	0.000836066
3360520	75909	Tmem49	3.309832056	0.000840952
5270603	106840	Unc119b	5.242387941	0.000840952
5260369	140546	Eri3	3.128622989	0.000844475
6580626	66588	Cmpk1	7.63058055	0.000848359
940435	11632	Aip	3.104322562	0.000860903
2490053	13636	Efna1	3.338370091	0.000860903
7560035	56449	Csda	3.167585408	0.000872308
3940494	214469	Fam168b	3.194386444	0.000890137
6100056	26399	Map2k6	3.060927923	0.000890137
6860390	140546	Eri3	5.863772404	0.000890137
1570100	21933	Tnfrsf10b	3.274334242	0.000890137
1030315	235527	Plscr4	3.308738087	0.000894844
2810040	219132	D14Ertd668e	3.122455292	0.000902937
6130551	26384	Gnpda1	4.370210232	0.000905553
4880356	20917	Suclg2	3.037327033	0.000915248
5890523	217430	Pqlc3	3.170311197	0.000916795
6110367	236848	BC023829	3.466228466	0.000950072
6420088	66540	Fam107b	3.394661898	0.000955661
5890544	28075	Pppde2	3.928512532	0.000955661
5670475	13002	Dnajc5	4.388238974	0.00098308
4880392	29871	Scmh1	3.490479913	0.000992355
3370487	14042	Ext1	3.517738499	0.001021786
2970091	66279	Tmem218	3.226807551	0.001034388
4610669	73910	Arhgap18	3.048529259	0.001038072
6580427	22193	Ube2e3	3.308501333	0.001038072
6330204	69860	Eif1ad	3.132198086	0.001054343
2230681	56443	Arpc1a	4.966834031	0.00106122
5570341	21374	Tbp	3.185560083	0.001066248
5690239	216344	Rab21	3.115978918	0.001075968
5490521	217718	Nek9	3.060732847	0.001089272
2900224	13207	Ddx5	3.176097598	0.001094388
60609	26413	Mapk1	2.994040365	0.00110034
3400270	69071	Tmem97	3.165461792	0.001114079
2850487	13664	Eif1a	2.997863212	0.001154185
2810025	54473	Tollip	2.912496159	0.001157499
5550647	72611	Zfp655	3.318549695	0.001157499
4540095	75616	2810008M24Rik	3.038663218	0.001172543

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
2070279	11637	Ak2	3.807129984	0.001232427
4220468	53379	Hnrnpa2b1	3.067213408	0.001242089
4610328	212427	A730008H23Rik	2.86226892	0.001252008
4010091	20615	Snapin	3.283153085	0.001252008
6020477	23938	Map2k5	2.860611062	0.001255124
2000291	225995	D030056L22Rik	3.154484197	0.001277347
4040341	56449	Csda	2.854800756	0.001278225
430215	11886	Asah1	2.912635217	0.001280125
4850373	231871	Daglb	4.068864638	0.001280125
6900408	70294	Rnf126	3.007430943	0.001288509
3420128	98710	Rabif	3.316361842	0.001332527
4070598	14391	Gabpb1	2.947564098	0.001336578
7200520	14628	Ostm1	2.913939624	0.001336578
50279	56422	Hbs1l	2.804373938	0.001377349
5700541	225027	Sfrs7	3.204509628	0.0013843
2690392	29857	Mapk12	5.681991239	0.001389186
6450546	51812	Mcrs1	2.81944432	0.001400104
4570414	29857	Mapk12	5.566793049	0.001403798
3780382	17129	Smad5	2.990533784	0.001403798
2490370	65257	Asb3	3.079647806	0.001417475
6580563	19298	Pex19	2.987425367	0.001417475
620010	75717	Cul5	2.993138356	0.001422705
3400349	56310	Gps2	2.992663812	0.001422705
4480068	217718	Nek9	2.913314197	0.001422705
6620209	19365	Rad52	2.826001734	0.001422705
3440725	20135	Rrm2	5.092085006	0.001422705
4850487	66827	Ttc1	3.115348295	0.001422705
5720202	12192	Zfp36l1	2.776681817	0.001422705
4250255	544752	Tug1	2.884755516	0.00142423
4560678	103534	Mgat4b	4.851961558	0.001434508
1660722	67840	Mrp63	2.797111576	0.001434508
5560646	20135	Rrm2	3.411038666	0.001434508
7560097	93697	Narg2	2.971128058	0.001445487
2120053	56443	Arpc1a	4.359366324	0.001489053
940180	216505	Pik3ip1	3.144004768	0.001523909
870121	70572	Ipo5	2.893791168	0.001526335
2750240	59008	Anapc5	5.790992204	0.001536318
2350706	66257	Nicn1	2.872761446	0.001536318
3780022	59054	Mrps30	2.731668099	0.001555323

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
4280445	66510	Rnf181	3.46551706	0.001596815
4900632	53902	Rcan3	2.743850071	0.001597954
1030296	56422	Hbs1l	2.719870609	0.001639753
5870561	107375	Slc25a45	2.851034031	0.001674594
870324	56233	Hdac7	2.81495238	0.001677626
2900402	26949	Vat1	3.266844299	0.001706578
3930202	76789	2410129H14Rik	3.673346577	0.001709364
1240300	26399	Map2k6	2.939274536	0.001711621
2850441	52830	Pnrc2	3.509745248	0.001725076
6660079	208638	Slc25a38	2.852390377	0.001725076
6520600	56708	Clcf1	3.736072531	0.001725961
3440242	72482	Acbd6	2.668385737	0.001734771
4280112	13360	Dhcr7	5.359144086	0.0017632
780711	171095	Il17rc	3.062793833	0.001766597
2120707	71801	Plekhf2	2.684074327	0.001766597
1110095	17828	Muted	2.68414376	0.001781832
1400685	230789	Fam76a	2.984138288	0.001800603
4860692	12400	Cbfb	2.896971917	0.001802404
6370270	105501	Abhd4	2.707617897	0.001803028
6110474	67921	Ube2f	3.041157263	0.001810136
5910601	70231	Gorasp2	3.021719691	0.001816934
650458	52118	Pvr	3.401778779	0.001834095
4060576	66674	6330409N04Rik	3.037793045	0.001848259
4570376	227358	Fam132b	2.648025461	0.001848259
6620609	66815	Ccdc109b	2.81811398	0.001851059
360626	14391	Gabpb1	2.626885029	0.001851059
1510270	83922	Tsga14	2.874164584	0.001858701
4390564	66612	Ormdl3	2.643469337	0.001898468
5490343	66257	Nicn1	2.781018886	0.001915885
4670047	68755	Cgrrf1	2.937314817	0.001958861
3830408	622320	Kctd21	2.738639601	0.001961803
3890609	14245	Lpin1	2.80805934	0.001961803
2350730	394430	Ugt1a10	2.634676609	0.001961803
6250681	217201	Rundc1	2.588197748	0.002035549
460474	66357	Ostc	2.952960866	0.00206069
3800537	234875	Ttc13	2.883500378	0.00206634
2070736	212127	2810046L04Rik	2.781780623	0.002075872
2900333	72599	Pdia5	2.576285393	0.002190327
6220202	56443	Arpc1a	3.378823328	0.002201438

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
2060435	20384	Sfrs5	2.733474489	0.002201438
2710019	214944	Mobkl2b	5.264709079	0.002221717
3420736	75734	Mff	2.560516634	0.002224002
6020605	19205	Ptbp1	2.560779526	0.002224002
1030170	26401	Map3k1	3.035959709	0.002237914
3840356	114714	Rad51c	2.550350894	0.002237914
2030739	22339	Vegfa	2.798588385	0.002243761
4050112	93692	Glrx	2.839484702	0.002264915
5960672	12400	Cbfb	3.144354071	0.002272516
580465	16599	Klf3	2.834909799	0.002278311
2230528	101739	Psip1	2.549775973	0.002278311
5820521	20250	Scd2	3.127191099	0.002284083
2000551	67916	Ppap2b	3.936700949	0.00232022
5960414	114863	Prosc	3.1524123	0.002323075
1170202	52477	Angel2	2.657698472	0.002336802
1570725	53902	Rcan3	2.59006666	0.002336802
3290541	239555	Smcr7l	3.146917042	0.002336802
1240678	19300	Abcd4	2.538920764	0.0023394
6400377	67861	Akr1b10	3.205969697	0.002356317
4060056	67695	2310016E02Rik	3.334901195	0.002356317
110594	99650	4933434E20Rik	2.644869602	0.002356317
4200347	56494	Gosr2	2.81924582	0.002356317
1660039	108052	Slc14a1	2.689189748	0.002356317
6760754	384864	Gm1943	3.205971474	0.002370471
430204	101739	Psip1	2.522140644	0.002370471
5560598	80294	Pofut2	3.490767025	0.002383646
270678	56468	Socs5	2.817298631	0.002383646
1980379	67846	Tmem39a	2.650441362	0.002383646
6860291	69674	Mif4gd	3.264578622	0.002404949
1010050	16911	Lmo4	2.667735406	0.00241607
1570014	76626	Msi2	2.754469461	0.00241607
6840121	18263	Odc1	2.620661283	0.00241607
2850703	24136	Zeb2	3.484990711	0.002437985
2760093	14950	H13	2.97397067	0.00248874
4540070	67878	Tmem33	2.568746721	0.002492997
2940681	14461	Gata2	2.604101551	0.002508895
6590445	245688	Rbbp7	3.060192597	0.002508895
670463	79560	Ublcp1	2.642653804	0.002508895
1990491	114713	Rasa2	3.06227144	0.002557144

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
6940255	67603	Dusp6	2.636407654	0.002579289
2450224	54197	Rnf5	2.624629565	0.002612116
1090100	21815	Tgif1	3.967751091	0.002617465
380240	72465	Zfp131	2.663606586	0.002661034
5890113	68606	Ppm1f	3.140175633	0.002722895
2320292	12499	Entpd5	2.446272301	0.002732139
3870768	232430	Crebl2	2.74529177	0.002746326
2370017	21933	Tnfrsf10b	2.515797972	0.002746326
1980168	228769	Psmf1	2.630323755	0.002771879
6960612	22196	Ube2i	3.267437765	0.002771879
1230722	78889	Wsb1	3.204942593	0.002798791
6760367	22628	Ywhag	2.436111281	0.002798791
4780373	19822	Rnf4	3.158762961	0.002826819
1510605	170750	Xpnpep1	3.111115444	0.002828319
270593	11804	Aplp2	3.809412925	0.002844239
3310181	54130	Actr1a	2.874943739	0.002874141
3130215	72050	Kdelc1	2.480087956	0.00298188
4920273	109129	Mmadhc	3.391673981	0.002986079
2760202	382406	Wdr51b	2.856111378	0.002986079
6060521	228876	Zfp334	2.453405605	0.003078834
160121	12908	Crat	2.447383142	0.003089508
4810577	78287	Zfyve20	2.398820842	0.003089508
1500373	72065	Rap2c	2.732261354	0.003194861
4570670	29819	Stau2	2.621190662	0.003194861
4280072	54208	Arl6ip1	2.528097362	0.003248626
6550343	12928	Crk	2.806916828	0.003248626
10176	319520	Dusp4	2.519163443	0.003248626
6520484	13690	Eif4g2	2.539150438	0.003248626
2070747	75612	Gns	2.659930371	0.003248626
7000465	68151	Gpr177	2.387588855	0.003248626
6060646	18515	Pbx2	2.808514975	0.003248626
670711	245688	Rbbp7	3.380775497	0.003261885
2680754	237082	Nxt2	2.517367714	0.003282901
3460670	67041	Oxct1	2.859343007	0.003282901
3800446	94284	Ugt1a6a	2.463475308	0.003331404
3800601	17828	Muted	2.441044873	0.003339412
3800274	20482	Skil	2.735743193	0.003403003
1230017	83396	Glis2	3.050940111	0.003406474
2340239	94065	Mrpl34	2.888086072	0.003406474

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
6020440	98910	Usp6nl	2.408818598	0.003406474
2650193	65098	Zfand6	3.453685796	0.003444891
5220187	74246	Gale	3.385753958	0.003446095
510341	50849	Rnf10	2.543418838	0.003451801
5960333	67921	Ube2f	3.000335872	0.003475472
2900592	14057	Sfxn1	2.796605221	0.00348087
1050379	140499	Ube2j2	2.483759851	0.00348087
3840762	74044	Ttf2	2.746884318	0.003544395
1440685	72881	Zdhhc4	3.67421067	0.003575505
7610161	71774	Shroom1	2.588556517	0.003677863
160064	76688	Arfrp1	2.433960379	0.003677863
2570132	17187	Max	2.600089815	0.003677863
2570754	81535	Sgpp1	2.466457336	0.003681613
1300372	13445	Cdk2ap1	2.585479344	0.003698024
1450164	67009	Ttc23	2.450389346	0.003699229
6330561	16784	Lamp2	2.353934299	0.003708007
780427	112405	Egln1	2.583428492	0.003715995
6040204	229658	Vangl1	2.509496837	0.003746657
650064	19345	Rab5c	2.374443182	0.003746657
3130048	19367	Rad9	2.514280747	0.003747938
1510020	19731	Rgl1	2.343069438	0.003747938
2750286	26384	Gnpda1	3.767653114	0.003755401
50136	320204	4833442J19Rik	2.365197547	0.003805555
2190743	71720	Osbpl3	2.322689101	0.003805555
5050634	80795	Selk	2.824102666	0.0038192
2100717	59038	Pxmp4	3.358680978	0.003845463
7560538	108052	Slc14a1	2.350074524	0.00386937
650411	17252	Rdh11	2.361946306	0.00389354
7560114	71720	Osbpl3	2.395967854	0.003899828
7650338	15980	Ifngr2	2.57735241	0.003949992
7330435	72465	Zfp131	2.423208093	0.003989067
6450056	20775	Sqle	2.323727466	0.004054224
7160343	14683	Gnas	2.732947875	0.004069549
1770435	70351	Ppp4r1	2.795718343	0.004069549
3830707	67958	2610101N10Rik	2.61911897	0.004077445
1780735	12125	Bcl2l11	2.642170297	0.004077445
670431	259300	Ehd2	2.650703735	0.004077445
2810280	71994	Cnn3	2.49294667	0.004077445
3170477	76281	Tax1bp3	2.365220468	0.004077445

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
1710746	13043	Cttn	2.428306759	0.004091038
520053	219132	D14Ertd668e	2.521366592	0.004091038
4290088	17169	Mark3	2.531407968	0.004098436
1410279	83767	Wasf1	2.496008587	0.004098436
6450291	330171	Kctd10	2.278746701	0.00416243
5080093	66471	Anp32e	2.390553314	0.004238939
5820672	21454	Tcp1	2.532959629	0.004253787
7380246	12972	Cryz	2.954481782	0.004258154
4890400	233724	Tmem41b	2.446966451	0.004258154
1580678	106957	Slc39a6	2.342199119	0.004270563
3870079	29857	Mapk12	2.478777194	0.00429975
2570747	53323	Ube2k	2.565943658	0.00429975
1710500	67771	Arpc5	2.4014256	0.004309855
6420520	14282	Fosb	2.618354216	0.004309855
4900747	21418	Tcfap2a	2.299255934	0.004323251
4570184	319765	Igf2bp2	2.683486492	0.004386669
5420458	67921	Ube2f	2.338664465	0.004410033
2070088	67149	Nkain1	2.445895888	0.004416216
1710193	14609	Gja1	2.303982929	0.004551183
6370025	108652	Slc35b3	2.320156333	0.004551183
1990541	21933	Tnfrsf10b	2.786959406	0.004610892
5960398	75541	1700019G17Rik	2.276997359	0.004614187
5080136	20405	Sh3gl1	2.386472312	0.004616525
4050768	56291	Styx	2.901547996	0.004627833
6840112	107652	Uap1	2.57809224	0.004627833
450491	18018	Nfatc1	2.327397847	0.004627835
4200468	107569	Nt5c3	2.56525316	0.004635887
4180086	214597	Sidt2	3.001068255	0.004635887
6130131	14950	H13	3.3277986	0.004642304
6200762	330064	Slc5a6	2.836807608	0.004676532
3450735	22375	Wars	2.403758578	0.004790759
130121	11886	Asah1	2.40212753	0.00480309
5870301	16975	Lrp8	2.231010029	0.00480309
5390669	14712	Gnpat	2.329061029	0.004816508
3140209	15945	Cxcl10	2.558193053	0.004834371
4920301	13852	Stx2	2.662009956	0.005025214
20400	234076	Tmco3	2.232352604	0.005025214
2060722	67921	Ube2f	2.369273832	0.005025214
2030286	244373	Erlin2	3.180796796	0.005031569

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
4880008	56298	Atl2	2.215224875	0.00512118
5910347	72033	Tsc22d2	2.283112076	0.005129591
6550133	75705	Eif4b	2.383052323	0.005130421
1690397	63873	Trpv4	3.391303892	0.005194622
4860255	20613	Snai1	2.270830559	0.005204941
4560114	50780	Rgs3	2.383975693	0.00527219
6940053	17344	Pias2	2.256379684	0.005281135
5260682	63873	Trpv4	3.291153257	0.005281135
160519	74868	Tmem65	2.476829945	0.005285967
1470554	14661	Glud1	2.64472164	0.005317515
4050328	66902	Mtap	2.191647298	0.005317515
2120048	59013	Hnrnph1	2.20601676	0.005331653
6290452	16526	Kcnk2	2.708539489	0.005350408
6550446	271457	Rab5a	2.532647169	0.005356835
2120201	99650	4933434E20Rik	2.530927846	0.005410487
50224	26394	Lypla2	2.32653999	0.005413026
1820224	17118	Marcks	2.229204234	0.005413026
4560309	18242	Oat	2.287496428	0.005413026
2750091	11566	Adss	2.357989602	0.005423731
160187	18602	Padi4	3.368660176	0.005442881
1070592	208638	Slc25a38	2.643390886	0.005471969
2120296	170750	Xpnpep1	2.56529258	0.005493028
2360131	72050	Kdelc1	2.255191415	0.005510821
5670192	15510	Hspd1	2.728588566	0.005515583
5130630	59003	Maea	3.303042788	0.005515583
5870537	109113	Uhrf2	2.832943586	0.005515583
3310402	232164	Paip2b	2.677116696	0.00556161
3180450	66366	Ergic3	3.828498026	0.005574107
2650598	245688	Rbbp7	2.897083877	0.005574107
7100326	17129	Smad5	2.263548372	0.005646738
110474	54128	Pmm2	2.589881346	0.005672602
510441	19367	Rad9	2.300258457	0.005687408
50603	23797	Akt3	2.455802902	0.00574416
2490209	76899	Golga1	2.529707195	0.00574416
7040719	68090	Yif1a	2.228253162	0.00574416
6040292	21807	Tsc22d1	3.752290459	0.005788052
360528	19707	Reps1	2.717747083	0.005827099
670164	56422	Hbs1l	2.270406132	0.00587121
3890687	81018	Rnf114	2.80806567	0.00587121

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
3780279	12028	Bax	2.294553037	0.005895339
6840433	18148	Npm1	2.928322271	0.005895339
3830114	14254	Flt1	2.461110354	0.005919982
450064	18537	Pcmt1	2.321726361	0.005923004
5960471	13929	Amz2	2.22861385	0.005992656
2480475	140742	Sesn1	2.354923583	0.006093557
4220605	228998	Arfgap1	2.479907855	0.006135465
2120475	23873	Faim	2.310172704	0.006178186
5290279	97165	Hmgb2	2.16828169	0.006262413
2510519	110842	Etfa	2.158692777	0.006262413
5130497	66286	Sec11c	2.361337405	0.006369503
1050382	54161	Copg	2.175791456	0.006372113
650168	117198	Ivns1abp	2.400912397	0.006372113
780338	223455	March6	2.404425	0.006372113
5340288	109232	Sccpdh	2.157824011	0.006372113
580717	109113	Uhrf2	2.277439237	0.006416578
3120446	27041	G3bp1	3.378433307	0.006446983
6900646	14009	Etv1	2.148460848	0.006457509
6220253	66493	Mrpl51	2.273343614	0.006457509
7000079	70351	Ppp4r1	2.229169914	0.006457509
5260201	60599	Trp53inp1	2.29460886	0.006459939
1710747	14536	Nr6a1	2.158212731	0.006487001
5890202	29864	Rnf11	2.141062452	0.006487001
2320189	14683	Gnas	3.863093513	0.006525917
6280484	67296	Socs4	2.146733668	0.006599571
1940156	22088	Tsg101	2.524334336	0.006613595
4900577	56708	Clcf1	3.161127733	0.006688864
4850039	52118	Pvr	2.211037783	0.006709221
1470598	76375	Det1	2.113047113	0.0067134
5270102	29864	Rnf11	2.122893397	0.0067134
5490403	76205	Stard3nl	2.715426492	0.006731797
1570543	15227	Foxf1a	2.470755975	0.006752698
3450241	23802	Amfr	2.105158952	0.006791792
3390735	50927	Nasp	2.178634893	0.006793053
4070019	214597	Sidt2	2.128107799	0.006807304
6220168	74197	Gtf2e1	2.177664587	0.007041038
7150541	11861	Arl4a	2.286731222	0.007042809
5700468	56205	Ensa	2.100463241	0.007088545
2630209	52118	Pvr	2.622428837	0.00710238

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
540288	19248	Ptpn12	3.09621306	0.007117421
6220750	68276	Toe1	2.159692724	0.007139644
1190239	66138	Wbscr22	2.691333374	0.007151053
2340438	208768	BC031781	2.16380928	0.007206172
10161	218581	Depdc1b	2.122236727	0.007261132
5890390	105513	Chmp7	2.776337111	0.007266321
3170414	108052	Slc14a1	2.120342996	0.007266321
5910554	73724	Mcee	2.681685422	0.007325284
6130632	15442	Hpse	2.118815151	0.007333941
1190112	15926	Idh1	2.225418293	0.007333941
6280376	214987	Chtf8	3.344019461	0.007444836
6620598	19042	Ppm1a	3.320300764	0.007455647
4290367	71279	Slc29a3	2.356697654	0.007522657
610164	75734	Mff	2.087808127	0.007577789
4540164	218335	Clptm1l	2.119309676	0.007740713
7510750	17129	Smad5	2.125892892	0.007743079
6350639	66049	Rogdi	2.145041821	0.007743079
3940017	94044	Bcl2l13	2.110635538	0.007785541
5700152	268470	Ube2z	2.409745331	0.007785541
4810468	14252	Flot2	2.146708961	0.007785541
5690347	71393	Kctd6	2.161597489	0.007785541
2470669	18516	Pbx3	2.104096081	0.007785541
6620717	72599	Pdia5	2.104334605	0.007785541
5820019	66817	Tmem170	2.184110724	0.007785541
3120259	81018	Rnf114	3.133699586	0.007785541
4150670	54613	St3gal6	2.087272002	0.007789235
6200092	11639	Ak3l1	2.326206907	0.007866019
3390121	225280	Ino80c	2.760319674	0.007880354
2320706	67223	Rrp15	2.476269115	0.007924328
2070615	11744	Anxa11	3.252968532	0.007958692
2000445	21379	Tbrg4	2.299575614	0.007958692
7050131	109910	Zfp91	2.161243086	0.007958692
4780475	232196	C87436	2.104637499	0.008019373
6520220	104318	Csnk1d	2.280447113	0.008098208
150128	53323	Ube2k	2.334934961	0.008120461
4760471	20218	Khdrbs1	2.23897466	0.008183375
510372	214597	Sidt2	2.053311712	0.008206903
670161	93684	Sep15	2.083605538	0.008294638
4570634	19328	Rab12	2.800427919	0.008327901

Table S2 (cont.). Set of genes enriched in the Bi-miR-200c PD with an enrichment ratio R>2 and P<0.01

ProbeID	Entrez_Gene_ID	Symbol	PD_ratio	Adj_P_value
3190048	320685	Dctd	2.056886481	0.008359108
4850528	72323	Asb6	2.475275999	0.008383468
540400	26426	Nubp2	2.346173531	0.008436514
7610059	263876	Spata2	2.386716602	0.008502206
6250092	53892	Ppm1d	2.049670125	0.00854654
3140411	234076	Tmco3	2.169403343	0.00860034
1820674	78308	Gpr108	2.240674979	0.008618662
6480333	60364	Donson	2.036249089	0.008712249
7160762	98758	Hnrnpf	2.595744953	0.008712249
3060162	18515	Pbx2	2.201822324	0.008733401
5870239	67105	1700034H14Rik	2.14289728	0.008911709
4010056	54635	Pdgfc	2.155758891	0.008940938
7510356	72999	Insig2	2.283055095	0.00917537
3390066	56462	Mtch1	2.836658032	0.009215223
6650376	21418	Tcfap2a	2.229584592	0.009271185
3450544	76178	6330578E17Rik	2.314055254	0.00928974
5720687	59013	Hnrnph1	2.043452856	0.00928974
1300164	21933	Tnfrsf10b	2.038248397	0.00928974
2970630	22147	Tuba3b	2.238512144	0.009312858
5050750	94112	Med15	2.312566286	0.00931926
3460468	20320	Nptn	2.296603972	0.00931926
6550392	106840	Unc119b	2.070523923	0.009377213
3170615	67270	D10Ertd322e	2.53784862	0.009419827
1440192	235442	Rab8b	2.246139708	0.009419827
7320239	53614	Reck	2.109142978	0.009429026
6900088	237339	L3mbtl3	2.209780038	0.00957953
780634	76281	Tax1bp3	2.10626494	0.009592018
3440187	68263	Pdhb	2.682085713	0.009641036
3710719	56190	Rbm38	3.464593959	0.009641036
2970424	69900	Mfsd11	2.274082374	0.009714707
4230019	66146	Tmem57	2.151041044	0.009714707
1110719	319710	Frmd6	2.106528464	0.009719436
4060180	72065	Rap2c	2.056454175	0.009820154
540475	52466	Slc46a1	2.020775825	0.009847382
4560424	13629	Eef2	4.381732169	0.009850651
1400402	76477	Pcolce2	2.175284807	0.009953539
7200181	227682	Trub2	2.209155352	0.009962634
3120491	80743	Vps16	2.991886615	0.009984913

Table S3. Bi-miR-200c PD genes (R>2, P<0.01) in each of the most enriched canonical pathways (Wikipathways) (AdjP<0.01)

Pathway	PD genes
MAPK signaling pathways	23797(Akt3) 12928(Crk) 12929(Crkl) 19252(Dusp1) 319520(Dusp4) 67603(Dusp6) 15481(Hspa8) 23938(Map2k5) 26399(Map2k6) 26401(Map3k1) 26413(Mapk1) 29857(Mapk12) 17187(Max) 17346(Mknk1) 19042(Ppm1a) 215449(Rap1b) 114713(Rasa2)
EGFR1 Signaling Pathway	12928(Crk) 12929(Crkl) 19252(Dusp1) 74155(Errfi1) 14609(Gja1) 13196(Asap1) 23938(Map2k5) 26401(Map3k1) 26413(Mapk1) 19248(Ptpn12) 271457(Rab5a) 245688(Rbbp7) 19707(Reps1) 17126(Smad2) 21815(Tgif1) 98910(Usp6nl) 54401(Ywhab)

Table S3 (cont.). Bi-miR-200c PD genes (R>2, P<0.01) in each of the most enriched canonical pathways (Wikipathways) (AdjP<0.01)

Pathway	PD genes
B Cell Receptor Signaling Pathway	56443(Arpc1a) 67771(Arpc5) 12028(Bax) 12042(Bcl10) 12125(Bcl2l11) 12928(Crk) 12929(Crkl) 319520(Dusp4) 67603(Dusp6) 56233(Hdac7) 26413(Mapk1) 18018(Nfatc1) 101476(Plekha1)
T Cell Receptor Signaling Pathway	12042(Bcl10) 12928(Crk) 12929(Crkl) 26431(Git2) 56233(Hdac7) 20218(Khdrbs1) 26401(Map3k1) 26413(Mapk1) 19248(Ptpn12) 192656(Ripk2)
Signal Transduction of S1P Receptor	23797(Akt3) 11886(Asah1) 26413(Mapk1) 29857(Mapk12)
Cholesterol Biosynthesis	13360(Dhcr7) 235293(Sc5d) 20775(Sqle)
Integrin-mediated cell adhesion	23797(Akt3) 12928(Crk) 26431(Git2) 23938(Map2k5) 26399(Map2k6) 26413(Mapk1) 29857(Mapk12) 215449(Rap1b)

Table S3 (cont.). Bi-miR-200c PD genes (R>2, P<0.01) in each of the most enriched canonical pathways (Wikipathways) (AdjP<0.01)

Pathway	PD genes
TGF beta signaling pathways	59008(Anapc5) 13559(E2f5) 14282(Fosb) 15481(Hspa8) 17129(Smad5) 26399(Map2k6) 17344(Pias2) 20482(Skil) 17126(Smad2) 21815(Tgif1) 24136(Zeb2)
Kit Receptor Signaling Pathway	12928(Crk) 12929(Crkl) 26413(Mapk1) 67296(Socs4) 56468(Socs5) 21933(Tnfrsf10b)
Signaling of Hepatocyte Growth Factor (HGF) Receptor	12928(Crk) 12929(Crkl) 26413(Mapk1) 215449(Rap1b)
IL-3 Signaling Pathway	12028(Bax) 12125(Bcl2l11) 12928(Crk) 12929(Crkl) 14461(Gata2) 26413(Mapk1) 54401(Ywhab)
Sphingolipid Metabolism	104725(Sptssa) 11886(Asah1) 56442(Serinc1) 22234(Ugcg)
Urea cycle and metabolism of amino groups	14661(Glud1) 18242(Oat) 18263(Odc1)

Table S4. Bi-miR-200c PD genes (R>2, P<0.001) in each of the most enriched canonical pathways (Wikipathways) (AdjP<0.05)

Pathway	PD genes
EGFR1 Signaling Pathway	12929(Crkl) 19252(Dusp1) 74155(Errfi1) 13196(Asap1) 271457(Rab5a) 17126(Smad2) 54401(Ywhab)
Cholesterol Biosynthesis	235293(Sc5d) 20775(Sqle)
Sphingolipid Metabolism	104725(Sptssa) 56442(Serinc1) 22234(Ugcg)
MAPK signaling pathways	12929(Crkl) 19252(Dusp1) 15481(Hspa8) 26399(Map2k6) 17346(Mknk1) 215449(Rap1b)
Diurnally regulated genes with circadian orthologs	15481(Hspa8) 18701(Pigf) 22057(Tob1)
Signaling of Hepatocyte Growth Factor Receptor	12929(Crkl) 215449(Rap1b)

Table S5. Predicted miR-200c MREs in the 3'UTRs of target genes (numbers in parentheses indicate the position of the MRE in each 3'UTR). Bottom sequence corresponds to miR-200c mature sequence. Experimental validation (Exp. validation) was performed by measuring luciferase reporter activity of constructs containing the wild-type or mutated seed residues (shown in red) 3'UTR of each gene. MREs that showed a statistically significant reduction in luciferase activity are defined as functional (see Figure 4). Perfect matches with miR-200c are indicated by a line; G:U pairs, by a colon.

3'UTR	Predicted MREs	Algorithm	Exp. validation
Crtap	MRE #1 (319-341) 5′ UCCCUGAAGAAAAAUCAGUAUUA 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (conserved MRE)	Functional
	MRE #2 (135-157) 5′ CCCCUUGUUCACACUCAGUAUUC 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Functional
Fhod1	MRE #1 (86-108) 5′ CUUAUACCCAGAAACCAGUAUUG 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Not functional
	MRE #2 (98-120) 5′ AACCAGUAUUGUGAACAGUAUUA 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (conserved MRE)	Functional
Tob1	MRE #1 (225-247) 5′ AGAUAUGGCCUCUUACAGUAUUU 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (conserved MRE)	Functional
Smad2	MRE #1 (287-309) 5′ UAAUGUGUCCAGACCCAGUAUUG 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Not functional
	MRE #2 (779-807) 5′ UCCAGGAAAGGUUUGUGUGGGCAGUAAGU 3′ 3′ AGGU----AGUAAUGG--GCCGUCAUAAU 5′	RNAhybrid	Functional
	MRE #3 (881-910) 5′ UGUCUCAAACAUGCACAGGUGGCAGUAUAA 3′ 3′ AGGUAG---UAAUG----GGCCGUCAUAAU 5′	RNAhybrid	Functional
Map3k1	MRE #1 (1865-1887) 5′ GGGACCUCAGCUAAUCAGUAUUA 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (conserved MRE)	Functional

Table S5 (cont.). Predicted miR-200c MREs in the 3'UTRs of target genes (numbers in parentheses indicate the position of the MRE in each 3'UTR). Bottom sequence corresponds to miR-200c mature sequence. Experimental validation (Exp. validation) was performed by measuring luciferase reporter activity of constructs containing the wild-type or mutated seed residues (shown in red) 3'UTR of each gene. MREs that showed a statistically significant reduction in luciferase activity are defined as functional (see Figure 4). Perfect matches with miR-200c are indicated by a line; G:U pairs, by a colon.

3'UTR	Predicted MREs	Algorithm	Exp. validation
Smad5	MRE #1 (3925-3947) 5′ UGCUAAGAAGGAUAACAGUAUUU 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Functional
Ywhag	MRE #1 (601-623) 5′ UCUCUGACGGGUUGACAGUAUUA 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (conserved MRE)	Not functional
	MRE #2 (627-649) 5′ CUCAAUUGUAGUUUACAGUAUUU 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Not tested
	MRE #3 (2005-2027) 5′ UCCAUCCUAUUGAUACAGUAUUG 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Functional
Xbp1	MRE #1 (444-466) 5′ AGAAGAAUUCCUCUAAAGUAUUU 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	PITA (MRE with lowest ΔΔG)	Functional
Zfp36	MRE #1 (581-603) 5′ UGACGACUUUAUUUAUU-GUAUUA 3′ 3′ AG-GUAGUAAUGGGCCGUCAUAAU 5′	RNAhybrid	Functional
	MRE #2 (658-680) 5′ AU----GUUAAAACUGAUACUGUAUUA 3′ 3′ AGGUAGUAAUG--GGCCGU--CAUAAU 5′	RNAhybrid	Not functional
Snail1	MRE #1 (181-203) 5′ GAGCCUCCUACCCCUCAGUAUUC 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Functional

Table S5 (cont.). Predicted miR-200c MREs in the 3'UTRs of target genes (numbers in parentheses indicate the position of the MRE in each 3'UTR). Bottom sequence corresponds to miR-200c mature sequence. Experimental validation (Exp. validation) was performed by measuring luciferase reporter activity of constructs containing the wild-type or mutated seed residues (shown in red) 3'UTR of each gene. MREs that showed a statistically significant reduction in luciferase activity are defined as functional (see Figure 4). Perfect matches with miR-200c are indicated by a line; G:U pairs, by a colon.

3'UTR	Predicted MREs	Algorithm	Exp. validation
Ywhab	MRE #1 (1640-1662) 5′ UGUUUUUUCCCUUACCAGUAUUC 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (conserved MRE)	Functional
Mapk12	MRE #1 (286-308) 5′ AGUUGUACAUGAAAUCAGUAUUC 3′ 3′ AGGUAGUAAUGGGCCGUCAUAAU 5′	TargetScan (poorly conserved MRE)	Functional

Table S6. List of qPCR and cloning primers.

	Primers used for qRT-PCR	
Gene	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Tuba1a	CTGGAACCCACGGTCATC	GTGGCCACGAGCATAGTTATT
Gapdh	GCACAGTCAAGGCCGAGAAT	GCCTTCTCCATGGTGGTGAA
Sdha	TGTTCAGTTCCACCCCACA	TCTCCACGACACCCTTCTG
Crtap	GAGCACTTCCAGCCCAGAC	TGTTCCTGAGCGAAGTCGTA
Fhod1	GCTCTGCACCGACTCCTG	ATCCGGAGGGAGACACTTG
Xbp1	TGACGAGGTTCCAGAGGTG	TGCAGAGGTGCACATAGTCTG
Zeb2	CAAGAGGCGCAAACAAGC	TGCGTCCACTACGTTGTCAT
Ywhag	CCCCAAACTTGCTTAACTTGTCT	GCTCAAGTGGGCGACAAA
Ywhab	GCAGGACATCTGCAACGA	CCTGGGTGGCATTGAGAAT
Ywhae	GATCGGGAGGATCTGGTGT	TCATTGATTCCACCATTTCG
Ywhah	ATTGGCAGCACAGCTATTCA	TGCCCATGAAGGTTTATCTGA
Ywhaq	CCCTCATCATGCAGTTGCT	TAGTTTTCGGCCCCCTCT
Ywhas	AAAGGGTGTGGCGAAGACTA	AGAAGAGTTGGGCGTGGTC
Ywhaz	CCAGACTGAGGAAGATTAAGCAAT	CAGGTTCCAGGTATCATTTGTAATTT
Zfp36	TCTCTTCACCAAGGCCATTC	ATCGACTGGAGGCTCTCG
Skp2	CGAAGCTTAGTCGGGAGAAC	AGGAGCAGCTCATCTGGAAG
Smad1	AACACCAGGCGACATATTGG	CACTGAGGCATTCCGCATA
Smad2	AGGACGGTTAGATGAGCTTGAG	GTCCCCAAATTTCAGAGCAA
Smad3	TCAAGAAGACGGGGCAGTT	CCGACCATCCAGTGACCT
Smad4	CCGTGGGTGGAATAGCTC	GGTCATCCACACCGATGC
Smad5	GCAGTAACATGATTCCTCAGACC	GCGACAGGCTGAACATCTCT
Smad8	CGGATGAGCTTTGTGAAGG	GGGTGCTCGTGACATCCT
Ajuba	AAAATTACGCCCCGAAATG	TCACCCTCACAATGTCCTCA
Prmt5	GCTGTCACCTGAGTGTCTGG	ATGCTCACGCCATCATCTTT
Map3k1	GCACATCCACATCTAGTTCAGAAA	CAGCAAGCAGATGGGACAC
Snai1	CTTGTGTCTGCACGACCTGT	AGGAGAATGGCTTCTCACCA
Scmh1	ACGGCGGTGAAGTCATTT	ACGGAGGACATAGGTGATGC
Mapk12	ACCTGATGAGTCTCTGGACGA	CCAGATCAGTGCCCATGAAT
Tob1	ACTTTTGCTGCCACCAAGTT	GAGCTACCTTGCTGCTACGG
Skil	TGCGTCCCAGTCTAAAGAGG	GCACACAGCAGACTCAGATTTC
Bcl2l11	GGAGACGAGTTCAACGAAACTT	AACAGTTGTAAGATAACCATTTGAGG
Mapk1	AAGAACTCATTTTTGAAGAGACTGC	CTCTGAGCCCTTGTCCTGA

Table S6 (cont.). List of qPCR and cloning primers.

	Primers used for qRT-PCR	
Gene	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Bax	GTGAGCGGCTGCTTGTCT	GGTCCCGAAGTAGGAGAGGA
Cdh1	ATCCTCGCCCTGCTGATT	ACCACCGTTCTCCTCCGTA
Cldn3	TGGGAGCTGGGTTGTACG	CAGGAGCAACACAGCAAGG
Cldn7	GACGCCCATGAACGTTAAGTA	CCTGGACAGGAGCAAGAGAG

	Primers used for ChIP assay	
	Sequence of Forward Primer (5' to 3')	**Sequence of Reverse Primer (5' to 3')**
Cdh1 promoter	CGAACGACCGTGGAATAGGA	CCTCCACCCCTGTCTGTAGT
Cdh1 exon	TTCCCAAGATCTTTCGTGGT	CAAGGAATGCCACTGTAAATGA
Cldn3 promoter	CTCTGCAGGTGGTCGTTGTC	AGTCCCTCAGGCTTTGACAC
Cldn3 exon	GTGGCCACTGCAGCTACTT	GTTTCATGGTTTGCCTGTCTC
Hbb1 promoter	GCCTTGCCTGTTCCTGCTC	ATTGAGCCCTTTACTCTCTCTGTTC

	Primers used for cloning 3'UTRs of miR-200c target genes	
3'UTR	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Zeb2	ATGCTCGAGACTACTGCATTTTAAGC TTCCTATTTT	ATGGCGGCCGCTGTTGCCAATTCTCTC CATGT
Crtap	ATGCTCGAGTCCACAGGGGCTAAGG AAC	ATGGCGGCCGCCCAAAGAAATTCTCCA AATATATTT
Fhod1	ATGCTCGAGGTGCTACCACGCCTGG AC	ATGGCGGCCGCTTAGCTAGGTAAACTT TATTTTGTTCC
Tob1	ATGCTCGAGAAAAACAAAACAAAGG AGAACATGT	ATGGCGGCCGCTTTTTAATTTTAAAGAA TTTTAATA
Smad2	ATGCTCGAGACCCATCAAAGACTCG CTGTAACAG	ATGGCGGCCGCCGGCTGCCTTTACAGA AAGGTACTT
Map3k1	ATGCTCGAGTTAATTGTTCAGATCAG CTCTAAT	ATGGCGGCCGCTTTTGTTCAAAAACTTT TATTTACT
Smad5	ATGGCGATCGCTGCAGAAGTATTCT TTCAACTATAT	ATGGCGGCCGCGTAAGCAGATTTTGAG GTTTATTAA
Ywhag	ATGCTCGAGAACAACTAAGGCCCCA GGT	ATGGCGGCCGCTCAGATCTGTAAGTTT ATTTTGCTC
Xbp1	ATGCTCGAGTTTTAGAGATCCCCTCT GAGACAGA	ATGGCGGCCGCTGCAAGGGTAGTTTTA AATAAATGG
Zfp36	ATGCTCGAGCAAGTGCCTACCTACC CAGTATGGA	ATGGCGGCCGCCAGTTCACACTCAAGA TTTTATTTA
Snai1	ATGCTCGAGCCCTGCTACCTCCCCA TC	ATGGCGGCCGCTAGTATATAAATTAACC ATTTTATTGAATATCAA

Table S6 (cont.). List of qPCR and cloning primers.

	Primers used for cloning 3'UTRs of miR-200c target genes	
3'UTR	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Ywhab	ATGCTCGAGTGTCTGTTGTGCTCTGTGATCT	ATGGCGGCCGCAACTGATAAAATGTTTTATTTCACTTT
Mapk12	ATGCTCGAGCGACCTCTGGGTGGTTTGGGGGGTA	ATGGCGGCCGCTGGGAGATCTACAAAAGGGTTTATT
Skp2	ATGCTCGAGGTATGATTTACATTTACGCTGGTGA	ATGGCGGCCGCTTGTTCTTATAAAAATGTATTTATT
Scmh1	ATGCTCGAGACAGGAGGCACTCTTCTCCCAGGAA	ATGGCGGCCGCTAAAAAGACACGGGCATTTGTATTG

	Primers used to mutate miR-200c MREs in 3'UTRs of target genes	
3'UTR	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Crtap MRE1_mut	GTCATAATTTTTTTAAGTAAGAAAAACACTAAAAGATG	ATTTTTCTTCAGGGATGAGAAAGC
Fhod1 MRE2_mut	GTCATAAAGGTGTGTGTGTGTGTGTGTG	TTCACAATACTGGTTTCTGGGTAT
Tob1 MRE1_mut	GTCATAATAAGATATAGCAAGGACATGG	TAAGAGGCCATATCTGAAAG
Smad2 MRE4_mut	CCGTCATAGTCCTTTTGTCTAATAAC	CACACAAACCTTTCCTGGAC
Smad2 MRE5_mut	ACCGTCATAAATTATTTTCAGGG	CCTGTGCATGTTTGAGACAAAAG
Map3k1 MRE1_mut	GTCATAATCCATGTAGATCAACATG	ATTAGCTGAGGTCCCTCAGTT
Smad5 MRE1_mut	GTCATAATAATAGTGTCTGTCATCAGG	TTATCCTTCTTAGCACACTAGAAC
Ywhag MRE3_mut	GTCATAAGTGCATGCTGGTGTTGTATTTC	TATCAATAGGATGGAGGAGCAG
Xbp1 MRE1_mut	TCATAATCAAATCTCATAGATGACTTCCAAG	TTAGAGGAATTCTTCTAAGGCCAGACAG
Zfp36 MRE4_mut	CATAATAGATTTTATAGTATTTATATATATTGG	AATAAATAAAGTCGTCATAAATAAAG
Snai1 MRE1_mut	GTCATAACATGAGGTGTAGCCTCTGGA	AGGGGTAGGAGGCTCTGG

Table S6 (cont.). List of qPCR and cloning primers.

	Primers used to mutate miR-200c MREs in 3'UTRs of target genes	
3'UTR	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Ywhab MRE1_mut	GTCATAACTGCCTAATGTTTGCTTCTGTGG	GTAAGGGAAAAAACAAAACAAACAAAC
Mapk12 MRE1_mut	GTCATAACGTGAAAAAGCTTCAGAGTGAG	ATTTCATGTACAACTCCAGAATTCAA

	Primers used to clone *Smad5* and *Ywhag* CDS	
CDS	**Sequence of Forward Primer (5' to 3')**	**Sequence of Reverse Primer (5' to 3')**
Smad5	ATGGCGGCCGCAAGATGACGTCAATGGCCAGCTTG	ATGCTCGAGCTATGAAACAGAAGAAATGGGG
Ywhag	ATGGCGGCCGCAAGATGGTGGACCGCGAGCAAC	ATGCTCGAGTTAGTTGTTGCCTTCACCGC

	siRNA sequences	
	sense	antisense
Firefly luciferase siRNA	5'P-CUUACGCUGAGUACUUCGAdTdT-3'	5'P-UCGAAGUACUCAGCGUAAGdTdT-3

9. References

1 Baum B, Settleman J, Quinlan MP. Transitions between epithelial and mesenchymal states in development and disease. *Semin Cell Dev Biol* 2008; **19:** 294-308.

2 Thiery JP. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* 2002; **2:** 442-454.

3 Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *J Clin Invest* 2009; **119:** 1420-1428.

4 Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F *et al*. miR-200 enhances mouse breast cancer cell colonization to form distant metastases. *PLoS One* 2009; **4:** e7181.

5 Wells A, Yates C, Shepard CR. E-cadherin as an indicator of mesenchymal to epithelial reverting transitions during the metastatic seeding of disseminated carcinomas. *Clin Exp Metastasis* 2008; **25:** 621-628.

6 Cheung KJ, Gabrielson E, Werb Z, Ewald AJ. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* 2013; **155:** 1639-1651.

7 Gravgaard KH, Lyng MB, Laenkholm AV, Sokilde R, Nielsen BS, Litman T *et al*. The miRNA-200 family and miRNA-9 exhibit differential expression in primary versus corresponding metastatic tissue in breast cancer. *Breast Cancer Res Treat* 2012; **134:** 207-217.

8 Korpal M, Ell BJ, Buffa FM, Ibrahim T, Blanco MA, Celia-Terrassa T *et al*. Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* 2011; **17:** 1101-1108.

9 Dai Y, Xia W, Song T, Su X, Li J, Li S *et al*. MicroRNA-200b Is Overexpressed in Endometrial Adenocarcinomas and Enhances MMP2 Activity by Downregulating TIMP2 in Human Endometrial Cancer Cell Line HEC-1A Cells. *Nucleic Acid Ther* 2013.

10 Li J, Du L, Yang Y, Wang C, Liu H, Wang L *et al*. MiR-429 is an independent prognostic factor in colorectal cancer and exerts its anti-apoptotic function by targeting SOX2. *Cancer Lett* 2013; **329:** 84-90.

11 Pichler M, Ress AL, Winter E, Stiegelbauer V, Karbiener M, Schwarzenbacher D *et al*. MiR-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. *Br J Cancer* 2014; **110:** 1614-1621.

12 Sun L, Yao Y, Liu B, Lin Z, Lin L, Yang M *et al*. MiR-200b and miR-15b regulate chemotherapy-induced epithelial-mesenchymal transition in human tongue cancer cells by targeting BMI1. *Oncogene* 2012; **31:** 432-445.

13 Ye F, Tang H, Liu Q, Xie X, Wu M, Liu X *et al*. miR-200b as a prognostic factor in breast cancer targets multiple members of RAB family. *J Transl Med* 2014; **12:** 17-27.

14 Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M *et al*. Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. *Cancer Res* 2010; **70:** 5226-5237.

15 Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL *et al*. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A* 2008; **105:** 10513-10518.

16 Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol* 2008; **110:** 13-21.

17 Toiyama Y, Hur K, Tanaka K, Inoue Y, Kusunoki M, Boland CR *et al*. Serum miR-200c Is a Novel Prognostic and Metastasis-Predictive Biomarker in Patients With Colorectal Cancer. *Ann Surg* 2014; **259:** 735-743.

18 Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S *et al*. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO Rep* 2008; **9:** 582-589.

19 Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* 2008; **283:** 14910-14914.

20 Park SM, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* 2008; **22:** 894-907.

21 Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G *et al*. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 2008; **10:** 593-601.

22 Chan YC, Khanna S, Roy S, Sen CK. miR-200b targets Ets-1 and is down-regulated by hypoxia to induce angiogenic response of endothelial cells. *J Biol Chem* 2011; **286:** 2047-2056.

23 Shin JO, Lee JM, Cho KW, Kwak S, Kwon HJ, Lee MJ *et al*. MiR-200b is involved in Tgf-beta signaling to regulate mammalian palate development. *Histochem Cell Biol* 2012; **137:** 67-78.

24 Liu YN, Yin JJ, Abou-Kheir W, Hynes PG, Casey OM, Fang L *et al*. MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. *Oncogene* 2013; **32:** 296-306.

25 Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D *et al*. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* 2009; **138:** 592-603.

26 Iliopoulos D, Lindahl-Allen M, Polytarchou C, Hirsch HA, Tsichlis PN, Struhl K. Loss of miR-200 Inhibition of Suz12 Leads to Polycomb-Mediated Repression Required for the Formation and Maintenance of Cancer Stem Cells. *Mol Cell* 2010; **39:** 761-772.

27 Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A *et al*. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 2010; **7:** 64-77.

28 Uhlmann S, Zhang JD, Schwager A, Mannsperger H, Riazalhosseini Y, Burmester S *et al*. miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene* 2010; **29:** 4297-4306.

29 Xia W, Li J, Chen L, Huang B, Li S, Yang G *et al*. MicroRNA-200b regulates cyclin D1 expression and promotes S-phase entry by targeting RND3 in HeLa cells. *Mol Cell Biochem* 2010; **344:** 261-266.

30 Yao CX, Wei QX, Zhang YY, Wang WP, Xue LX, Yang F *et al*. miR-200b targets GATA-4 during cell growth and differentiation. *RNA Biol* 2013; **10:** 465-480.

31 Schickel R, Park SM, Murmann AE, Peter ME. mir-200c Regulates Induction of Apoptosis through CD95 by Targeting FAP-1. *Mol Cell* 2010; **38:** 908-915.

32 Feng B, Wang R, Song HZ, Chen LB. MicroRNA-200b reverses chemoresistance of docetaxel-resistant human lung adenocarcinoma cells by targeting E2F3. *Cancer* 2012; **118:** 3365-3376.

33 Zhu W, Xu H, Zhu D, Zhi H, Wang T, Wang J *et al*. miR-200bc/429 cluster modulates multidrug resistance of human cancer cell lines by targeting BCL2 and XIAP. *Cancer Chemother Pharmacol* 2012; **69:** 723-731.

34 Lal A, Thomas MP, Altschuler G, Navarro F, O'Day E, Li XL *et al*. Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet* 2011; **7:** e1002363.

35 Orom UA, Nielsen FC, Lund AH. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* 2008; **30:** 460-471.

36 Krishnan K, Steptoe AL, Martin HC, Wani S, Nones K, Waddell N *et al*. MicroRNA-182-5p targets a network of genes involved in DNA repair. *RNA* 2013; **19:** 230-242.

37 Kang H, Davis-Dusenbery BN, Nguyen PH, Lal A, Lieberman J, Van Aelst L *et al*. Bone morphogenetic protein 4 promotes vascular smooth muscle contractility by activating microRNA-21 (miR-21), which down-regulates expression of family of dedicator of cytokinesis (DOCK) proteins. *J Biol Chem* 2012; **287:** 3976-3986.

38 Cloonan N, Wani S, Xu Q, Gu J, Lea K, Heater S *et al*. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol* 2011; **12:** R126.

39 Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W *et al*. Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. *Cell Rep* 2014; **8:** 1225-1239.

40 Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F *et al*. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* 2013; **154:** 61-74.

41 Verschueren K, Remacle JE, Collart C, Kraft H, Baker BS, Tylzanowski P *et al*. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5'-CACCT sequences in candidate target genes. *J Biol Chem* 1999; **274:** 20489-20498.

42 Hou Z, Peng H, White DE, Wang P, Lieberman PM, Halazonetis T *et al*. 14-3-3 binding sites in the snail protein are essential for snail-mediated transcriptional repression and epithelial-mesenchymal differentiation. *Cancer Res* 2010; **70:** 4385-4393.

43 Aslakson CJ, Miller FR. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res* 1992; **52:** 1399-1405.

44 Petrocca F, Altschuler G, Tan SM, Mendillo ML, Yan H, Jerry DJ *et al*. A genome-wide siRNA screen identifies proteasome addiction as a vulnerability of basal-like triple-negative breast cancer cells. *Cancer Cell* 2013; **24:** 182-196.

45 Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH *et al*. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol Cell* 2007; **26:** 745-752.

46 Chen R, Li L, Butte AJ. AILUN: reannotating gene expression data automatically. *Nat Methods* 2007; **4:** 879.

47 Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; **120:** 15-20.

48 Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 2007; **27:** 91-105.

49 Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. *Nat Struct Mol Biol* 2011; **18:** 1139-1146.

50 Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. *Nat Genet* 2007; **39:** 1278-1284.

51 Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA* 2004; **10:** 1507-1517.

52 Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol* 2008; **6:** e184.

53 Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J *et al*. CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* 2012; **72:** 3499-3511.

54 Obri A, Ouararhni K, Papin C, Diebold ML, Padmanabhan K, Marek M *et al*. ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature* 2014; **505:** 648-653.

55 Tanjore H, Cheng DS, Degryse AL, Zoz DF, Abdolrasulnia R, Lawson WE *et al*. Alveolar Epithelial Cells Undergo Epithelial-to-Mesenchymal Transition in Response to Endoplasmic Reticulum Stress. *J Biol Chem* 2011; **286:** 30972-30980.

56 Chen X, Iliopoulos D, Zhang Q, Tang Q, Greenblatt MB, Hatziapostolou M *et al*. XBP1 promotes triple-negative breast cancer by controlling the HIF1alpha pathway. *Nature* 2014; **508:** 103-107.

57 Koka S, Neudauer CL, Li X, Lewis RE, McCarthy JB, Westendorf JJ. The formin-homology-domain-containing protein FHOD1 enhances cell migration. *J Cell Sci* 2003; **116:** 1745-1755.

58 Yamazaki D, Fujiwara T, Suetsugu S, Takenawa T. A novel function of WAVE in lamellipodia: WAVE1 is required for stabilization of lamellipodial protrusions during cell spreading. *Genes Cells* 2005; **10:** 381-392.

59 Cuevas BD, Abell AN, Witowsky JA, Yujiri T, Johnson NL, Kesavan K *et al*. MEKK1 regulates calpain-dependent proteolysis of focal adhesion proteins for rear-end detachment of migrating fibroblasts. *EMBO J* 2003; **22:** 3346-3355.

60 Comijn J, Berx G, Vermassen P, Verschueren K, van Grunsven L, Bruyneel E *et al*. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol Cell* 2001; **7:** 1267-1278.

61 Batlle E, Sancho E, Franci C, Dominguez D, Monfar M, Baulida J *et al*. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat Cell Biol* 2000; **2:** 84-89.

62 Liu IM, Schilling SH, Knouse KA, Choy L, Derynck R, Wang XF. TGFbeta-stimulated Smad1/5 phosphorylation requires the ALK5 L45 loop and mediates the pro-migratory TGFbeta switch. *EMBO J* 2009; **28:** 88-98.

63 Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG *et al*. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* 2011; **6:** e16915.

64 Garibaldi F, Cicchini C, Conigliaro A, Santangelo L, Cozzolino AM, Grassi G *et al*. An epistatic mini-circuitry between the transcription factors Snail and HNF4alpha controls liver stem cell and hepatocyte features exhorting opposite regulation on stemness-inhibiting microRNAs. *Cell Death Differ* 2012; **19:** 937-946.

65 Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF *et al*. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res* 2008; **68:** 7846-7854.

66 Itoh T, Nozawa Y, Akao Y. MicroRNA-141 and -200a are involved in bone morphogenetic protein-2-induced mouse pre-osteoblast differentiation by targeting distal-less homeobox 5. *J Biol Chem* 2009; **284:** 19272-19279.

67 Gregory PA, Bracken CP, Smith E, Bert AG, Wright JA, Roslan S *et al*. An autocrine TGF-beta/ZEB/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol Biol Cell* 2011; **22:** 1686-1698.

68 Zhou BP, Deng J, Xia W, Xu J, Li YM, Gunduz M *et al*. Dual regulation of Snail by GSK-3beta-mediated phosphorylation in control of epithelial-mesenchymal transition. *Nat Cell Biol* 2004; **6:** 931-940.

69 Langer EM, Feng Y, Zhaoyuan H, Rauscher FJ, 3rd, Kroll KL, Longmore GD. Ajuba LIM proteins are snail/slug corepressors required for neural crest development in Xenopus. *Dev Cell* 2008; **14:** 424-436.

70 Venkov CD, Link AJ, Jennings JL, Plieth D, Inoue T, Nagai K *et al*. A proximal activator of transcription in epithelial-mesenchymal transition. *J Clin Invest* 2007; **117:** 482-491.

71 Mishra SK, Talukder AH, Gururaj AE, Yang Z, Singh RR, Mahoney MG *et al*. Upstream determinants of estrogen receptor-alpha regulation of metastatic tumor antigen 3 pathway. *J Biol Chem* 2004; **279:** 32709-32715.

72 Ahmadi H, Ahmadi A, Azimzadeh-Jamalkandi S, Shoorehdeli MA, Salehzadeh-Yazdi A, Bidkhori G *et al*. HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens. *Genomics* 2013; **101:** 94-100.

73 Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009; **460:** 479-486.

74 Jurmeister S, Baumann M, Balwierz A, Keklikoglou I, Ward A, Uhlmann S *et al*. MicroRNA-200c represses migration and invasion of breast cancer cells by targeting actin-regulatory proteins FHOD1 and PPM1F. *Mol Cell Biol* 2012; **32:** 633-651.

75 Capobianco V, Nardelli C, Ferrigno M, Iaffaldano L, Pilone V, Forestieri P *et al*. miRNA and Protein Expression Profiles of Visceral Adipose Tissue Reveal miR-141/YWHAG and miR-520e/RAB11A as Two Potential miRNA/Protein Target Pairs Associated with Severe Obesity. *J Proteome Res* 2012.

76 Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* 2013; **153:** 654-665.

77 Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P *et al*. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 2010; **141:** 129-141.

78 Grafe I, Yang T, Alexander S, Homan EP, Lietman C, Jiang MM *et al*. Excessive transforming growth factor-beta signaling is a common mechanism in osteogenesis imperfecta. *Nat Med* 2014; **20:** 670-675.

79 Bracken CP, Li X, Wright JA, Lawrence DM, Pillman KA, Salmanidis M *et al*. Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. *EMBO J* 2014; **33:** 2040-2056.

80 Senanayake U, Das S, Vesely P, Alzoughbi W, Frohlich LF, Chowdhury P *et al*. miR-192, miR-194, miR-215, miR-200c and miR-141 are downregulated and their common target ACVR2B is strongly expressed in renal childhood neoplasms. *Carcinogenesis* 2012; **33:** 1014-1021.

81 Wu Y, Xiao Y, Ding X, Zhuo Y, Ren P, Zhou C *et al*. A miR-200b/200c/429-binding site polymorphism in the 3' untranslated region of the AP-2alpha gene is associated with cisplatin resistance. *PLoS One* 2011; **6:** e29043.

82 Howe EN, Cochrane DR, Richer JK. Targets of miR-200c mediate suppression of cell motility and anoikis resistance. *Breast Cancer Res* 2011; **13:** R45.

83 Wang Y, Li M, Zang W, Ma Y, Wang N, Li P *et al*. MiR-429 up-regulation induces apoptosis and suppresses invasion by targeting Bcl-2 and SP-1 in esophageal carcinoma. *Cell Oncol (Dordr)* 2013; **36:** 385-394.

84 Yin J, Zheng G, Jia X, Zhang Z, Zhang W, Song Y *et al*. A Bmi1-miRNAs cross-talk modulates chemotherapy response to 5-fluorouracil in breast cancer cells. *PLoS One* 2013; **8:** e73268.

85 Lo WL, Yu CC, Chiou GY, Chen YW, Huang PI, Chien CS *et al*. MicroRNA-200c attenuates tumour growth and metastasis of presumptive head and neck squamous cell carcinoma stem cells. *J Pathol* 2011; **223:** 482-495.

86 Wellner U, Schubert J, Burk UC, Schmalhofer O, Zhu F, Sonntag A *et al*. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol* 2009; **11:** 1487-1495.

87 Dimri M, Carroll JD, Cho JH, Dimri GP. microRNA-141 regulates BMI1 expression and induces senescence in human diploid fibroblasts. *Cell Cycle* 2013; **12:** 3537-3546.

88 Cui J, Cheng Y, Zhang P, Sun M, Gao F, Liu C *et al*. Down Regulation of miR200c Promotes Radiation-Induced Thymic Lymphoma by Targeting BMI1. *J Cell Biochem* 2013; **115:** 1033-1042.

89 Zhang L, Deng T, Li X, Liu H, Zhou H, Ma J *et al*. microRNA-141 is involved in a nasopharyngeal carcinoma-related genes network. *Carcinogenesis* 2010; **31:** 559-566.

90 Park YA, Lee JW, Choi JJ, Jeon HK, Cho Y, Choi C *et al*. The interactions between MicroRNA-200c and BRD7 in endometrial carcinoma. *Gynecol Oncol* 2012; **124:** 125-133.

91 Kim Y, Park D, Kim H, Choi M, Lee H, Lee YS *et al*. miR-200b and cancer/testis antigen CAGE form a feedback loop to regulate the invasion and tumorigenic and angiogenic responses of a cancer cell line to microtubule-targeting drugs. *J Biol Chem* 2013; **288:** 36502-36518.

92 Tang W, Qin J, Tang J, Zhang H, Zhou Z, Li B *et al*. Aberrant Reduction of MiR-141 Increased CD47/CUL3 in Hirschsprung's Disease. *Cell Physiol Biochem* 2013; **32:** 1655-1667.

93 Yu XY, Zhang Z, Liu J, Zhan B, Kong CZ. MicroRNA-141 is downregulated in human renal cell carcinoma and regulates cell survival by targeting CDC25B. *Onco Targets Ther* 2013; **6:** 349-354.

94 Luo D, Wilson JM, Harvel N, Liu J, Pei L, Huang S *et al*. A systematic evaluation of miRNA:mRNA interactions involved in the migration and invasion of breast cancer cells. *J Transl Med* 2013; **11:** 57.

95 Lin CH, Jackson AL, Guo J, Linsley PS, Eisenman RN. Myc-regulated microRNAs attenuate embryonic stem cell differentiation. *EMBO J* 2009; **28:** 3157-3170.

96 Zhang HF, Zhang K, Liao LD, Li LY, Du ZP, Wu BL *et al*. miR-200b suppresses invasiveness and modulates the cytoskeletal and adhesive machinery in esophageal squamous cell carcinoma cells via targeting Kindlin-2. *Carcinogenesis* 2014; **35:** 292-301.

97 Xiao F, Zhang W, Zhou L, Xie H, Xing C, Ding S *et al*. microRNA-200a is an independent prognostic factor of hepatocellular carcinoma and induces cell cycle arrest by targeting CDK6. *Oncol Rep* 2010; **30:** 2203-2210.

98 Fu Y, Liu X, Zhou N, Du L, Sun Y, Zhang X *et al*. MicroRNA-200b stimulates tumour growth in TGFBR2-null colorectal cancers by negatively regulating p27/kip1. *J Cell Physiol* 2013; **229:** 772-782.

99 Klein D, Misawa R, Bravo-Egana V, Vargas N, Rosero S, Piroso J *et al*. MicroRNA Expression in Alpha and Beta Cells of Human Pancreatic Islets. *PLoS One* 2013; **8:** e55064.

100 Cesi V, Casciati A, Sesti F, Tanno B, Calabretta B, Raschella G. TGFbeta-induced c-Myb affects the expression of EMT-associated genes and promotes invasion of ER+ breast cancer cells. *Cell Cycle* 2011; **10:** 4149-4161.

101 Sun T, Wang C, Xing J, Wu D. miR-429 modulates the expression of c-myc in human gastric carcinoma cells. *Eur J Cancer* 2011; **47:** 2552-2559.

102 Peng B, Hu S, Jun Q, Luo D, Zhang X, Zhao H *et al*. MicroRNA-200b targets CREB1 and suppresses cell growth in human malignant glioma. *Mol Cell Biochem* 2013; **379:** 51-58.

103 Wu X, Zhao B, Li W, Chen Y, Liang R, Li L *et al*. MiR-200a is involved in rat epididymal development by targeting beta-catenin mRNA. *Acta Biochim Biophys Sin (Shanghai)* 2012; **44:** 233-240.

104 Sun X, He Y, Ma TT, Huang C, Zhang L, Li J. Participation of miR-200a in TGF-beta1-mediated hepatic stellate cell activation. *Mol Cell Biochem* 2014; **388:** 11-23.

105 Liu J, Ruan B, You N, Huang Q, Liu W, Dang Z *et al*. Downregulation of miR-200a induces EMT phenotypes and CSC-like signatures through targeting the beta-catenin pathway in hepatic oval cells. *PLoS One* 2013; **8:** e79409.

106 Saydam O, Shen Y, Wurdinger T, Senol O, Boke E, James MF *et al*. Downregulated microRNA-200a in meningiomas promotes tumor growth by reducing E-cadherin and activating the Wnt/beta-catenin signaling pathway. *Mol Cell Biol* 2009; **29:** 5923-5940.

107 Su J, Zhang A, Shi Z, Ma F, Pu P, Wang T *et al*. MicroRNA-200a suppresses the Wnt/beta-catenin signaling pathway by interacting with beta-catenin. *Int J Oncol* 2012; **40:** 1162-1170.

108 Pecot CV, Rupaimoole R, Yang D, Akbani R, Ivan C, Lu C *et al*. Tumour angiogenesis regulation by the miR-200 family. *Nat Commun* 2013; **4:** 2427.

109 Huang Z, Shi T, Zhou Q, Shi S, Zhao R, Shi H *et al*. miR-141 Regulates colonic leukocytic trafficking by targeting CXCL12beta during murine colitis and human Crohn's disease. *Gut* 2013.

110 Cai J, Liu X, Cheng J, Li Y, Huang X, Li Y *et al*. MicroRNA-200 is commonly repressed in conjunctival MALT lymphoma, and targets cyclin E2. *Graefes Arch Clin Exp Ophthalmol* 2012; **250:** 523-531.

111 Banaudha K, Kaliszewski M, Korolnek T, Florea L, Yeung ML, Jeang KT *et al*. MicroRNA silencing of tumor suppressor DLC-1 promotes efficient hepatitis C virus replication in primary human hepatocytes. *Hepatology* 2011; **53:** 53-61.

112 Tang H, Deng M, Tang Y, Xie X, Guo J, Kong Y *et al*. miR-200b and miR-200c as Prognostic Factors and Mediators of Gastric Cancer Cell Progression. *Clin Cancer Res* 2013; **19:** 5602-5612.

113 Peng C, Li N, Ng YK, Zhang J, Meier F, Theis FJ *et al*. A unilateral negative feedback loop between miR-200 microRNAs and Sox2/E2F3 controls neural progenitor cell-cycle exit and differentiation. *J Neurosci* 2012; **32:** 13292-13308.

114 Ho BC, Yu SL, Chen JJ, Chang SY, Yan BS, Hong QS *et al*. Enterovirus-induced miR-141 contributes to shutoff of host protein translation by targeting the translation initiation factor eIF4E. *Cell Host Microbe* 2011; **9:** 58-69.

115 Sun Q, Zou X, Zhang T, Shen J, Yin Y, Xiang J. The role of miR-200a in vasculogenic mimicry and its clinical significance in ovarian cancer. *Gynecol Oncol* 2014; **132:** 730-738.

116 Li Y, Nie Y, Cao J, Tu S, Lin Y, Du Y *et al*. G-A variant in miR-200c binding site of EFNA1 alters susceptibility to gastric cancer. *Mol Carcinog* 2012; **53:** 219-229.

117 Adam L, Zhong M, Choi W, Qi W, Nicoloso M, Arora A *et al*. miR-200 expression regulates epithelial-to-mesenchymal transition in bladder cancer cells and reverses resistance to epidermal growth factor receptor therapy. *Clin Cancer Res* 2009; **15:** 5060-5072.

118 Luna C, Li G, Huang J, Qiu J, Wu J, Yuan F *et al*. Regulation of trabecular meshwork cell contraction and intraocular pressure by miR-200c. *PLoS One* 2012; **7:** e51688.

119 Gill JG, Langer EM, Lindsley RC, Cai M, Murphy TL, Murphy KM. Snail promotes the cell-autonomous generation of Flk1(+) endothelial cells through the repression of the microRNA-200 family. *Stem Cells Dev* 2012; **21:** 167-176.

120 Chuang TD, Panda H, Luo X, Chegini N. miR-200c is aberrantly expressed in leiomyomas in an ethnic-dependent manner and targets ZEBs, VEGFA, TIMP2, and FBLN5. *Endocr Relat Cancer* 2012; **19:** 541-556.

121 Chan YC, Roy S, Khanna S, Sen CK. Downregulation of endothelial microRNA-200b supports cutaneous wound angiogenesis by desilencing GATA binding protein 2 and vascular endothelial growth factor receptor 2. *Arterioscler Thromb Vasc Biol* 2012; **32:** 1372-1382.

122 Choi YC, Yoon S, Jeong Y, Yoon J, Baek K. Regulation of vascular endothelial growth factor signaling by miR-200b. *Mol Cells* 2011; **32:** 77-82.

123 Shi L, Zhang S, Wu H, Zhang L, Dai X, Hu J *et al*. MiR-200c increases the radiosensitivity of non-small-cell lung cancer cell line A549 by targeting VEGF-VEGFR2 pathway. *PLoS One* 2013; **8:** e78344.

124 Panda H, Pelakh L, Chuang TD, Luo X, Bukulmez O, Chegini N. Endometrial miR-200c is altered during transformation into cancerous states and targets the expression of ZEBs, VEGFA, FLT1, IKKbeta, KLF9, and FBLN5. *Reprod Sci* 2012; **19:** 786-796.

125 Mezquita B, Mezquita J, Barrot C, Carvajal S, Pau M, Mezquita P *et al*. A truncated-Flt1 isoform of breast cancer cells is upregulated by Notch and downregulated by retinoic acid. *J Cell Biochem* 2014; **115:** 52-61.

126 Wei J, Zhang Y, Luo Y, Wang Z, Bi S, Song D *et al*. Aldose reductase regulates miR-200a-3p/141-3p to coordinate Keap1-Nrf2, Tgfbeta1/2, and Zeb1/2 signaling in renal mesangial cells and the renal cortex of diabetic mice. *Free Radic Biol Med* 2014; **67:** 91-102.

127 Tang O, Chen XM, Shen S, Hahn M, Pollock CA. MiRNA-200b represses transforming growth factor-beta1-induced EMT and fibronectin expression in kidney proximal tubular cells. *Am J Physiol Renal Physiol* 2013; **304:** F1266-1273.

128 Hyun S, Lee JH, Jin H, Nam J, Namkoong B, Lee G *et al*. Conserved MicroRNA miR-8/miR-200 and its target USH/FOG2 control growth by regulating PI3K. *Cell* 2009; **139:** 1096-1108.

129 Shpyleva SI, Tryndyak VP, Kovalchuk O, Starlard-Davenport A, Chekhun VF, Beland FA *et al*. Role of ferritin alterations in human breast cancer cells. *Breast Cancer Res Treat* 2011; **126:** 63-71.

130 Huang HN, Chen SY, Hwang SM, Yu CC, Su MW, Mai W *et al*. miR-200c and GATA binding protein 4 regulate human embryonic stem cell renewal and differentiation. *Stem Cell Res* 2013; **12:** 338-353.

131 Rasheed SA, Teo CR, Beillard EJ, Voorhoeve PM, Casey PJ. MicroRNA-182 and microRNA-200a control G-protein subunit alpha-13 (GNA13) expression and cell invasion synergistically in prostate cancer cells. *J Biol Chem* 2013; **288:** 7986-7995.

132 Liu Y, Liu Q, Jia W, Chen J, Wang J, Ye D *et al*. MicroRNA-200a regulates Grb2 and suppresses differentiation of mouse embryonic stem cells into endoderm and mesoderm. *PLoS One* 2013; **8:** e68990.

133 Yuan JH, Yang F, Chen BF, Lu Z, Huo XS, Zhou WP *et al*. The histone deacetylase 4/SP1/microrna-200a regulatory network contributes to aberrant histone acetylation in hepatocellular carcinoma. *Hepatology* 2011; **54:** 2025-2035.

134 Chen B, Huang T, Jiang J, Lv L, Li H, Xia S. miR-141 suppresses proliferation and motility of gastric cancer cells by targeting HDGF. *Mol Cell Biochem* 2014; **388:** 211-218.

135 Vallejo DM, Caparros E, Dominguez M. Targeting Notch signalling by the conserved miR-8/200 microRNA family in development and cancer cells. *EMBO J* 2011; **30:** 756-769.

136 Brabletz S, Bajdak K, Meidhof S, Burk U, Niedermann G, Firat E *et al*. The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. *EMBO J* 2011; **30:** 770-782.

137 Jia Y, Yang Y, Zhan Q, Brock MV, Zheng X, Yu Y *et al*. Inhibition of SOX17 by microRNA 141 and methylation activates the WNT signaling pathway in esophageal cancer. *J Mol Diagn* 2012; **14:** 577-585.

138 van Jaarsveld MT, Helleman J, Boersma AW, van Kuijk PF, van Ijcken WF, Despierre E *et al*. miR-141 regulates KEAP1 and modulates cisplatin sensitivity in ovarian cancer cells. *Oncogene* 2013; **32:** 4284-4293.

139 Eades G, Yang M, Yao Y, Zhang Y, Zhou Q. miR-200a regulates Nrf2 activation by targeting Keap1 mRNA in breast cancer cells. *J Biol Chem* 2011; **286:** 40725-40733.

140 Kopp F, Wagner E, Roidl A. The proto-oncogene KRAS is targeted by miR-200c. *Oncotarget* 2014; **5:** 185-195.

141 Zhao G, Wang B, Liu Y, Zhang JG, Deng SC, Qin Q *et al*. miRNA-141, downregulated in pancreatic cancer, inhibits cell proliferation and invasion by directly targeting MAP4K4. *Mol Cancer Ther* 2013; **12:** 2569-2580.

142 Mateescu B, Batista L, Cardon M, Gruosso T, de Feraudy Y, Mariani O *et al*. miR-141 and miR-200a act on ovarian tumorigenesis by controlling oxidative stress response. *Nat Med* 2011; **17:** 1627-1635.

143 Elson-Schwab I, Lorentzen A, Marshall CJ. MicroRNA-200 family members differentially regulate morphological plasticity and mode of melanoma cell invasion. *PLoS One* 2010; **5**.

144 Marasa BS, Srikantan S, Masuda K, Abdelmohsen K, Kuwano Y, Yang X *et al*. Increased MKK4 abundance with replicative senescence is linked to the joint reduction of multiple microRNAs. *Sci Signal* 2009; **2:** ra69.

145 Li X, Roslan S, Johnstone CN, Wright JA, Bracken CP, Anderson M *et al*. MiR-200 can repress breast cancer metastasis through ZEB1-independent but moesin-dependent pathways. *Oncogene* 2013.

146 Radhakrishnan P, Mohr AM, Grandgenett PM, Steele MM, Batra SK, Hollingsworth MA. MicroRNA-200c modulates the expression of MUC4 and MUC16 by directly targeting their coding sequences in human pancreatic cancer. *PLoS One* 2013; **8:** e73356.

147 Oishi N, Kumar MR, Roessler S, Ji J, Forgues M, Budhu A *et al*. Transcriptomic profiling reveals hepatic stem-like gene signatures and interplay of miR-200c and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma. *Hepatology* 2012; **56:** 1792-1803.

148 Cao H, Jheon A, Li X, Sun Z, Wang J, Florez S *et al*. The Pitx2:miR-200c/141:noggin pathway regulates Bmp signaling and ameloblast differentiation. *Development* 2013; **140:** 3348-3359.

149 Cama A, Verginelli F, Lotti LV, Napolitano F, Morgano A, D'Orazio A *et al*. Integrative genetic, epigenetic and pathological analysis of paraganglioma reveals complex dysregulation of NOTCH signaling. *Acta Neuropathol* 2013; **126:** 575-594.

150 Yang X, Ni W, Lei K. miR-200b suppresses cell growth, migration and invasion by targeting Notch1 in nasopharyngeal carcinoma. *Cell Physiol Biochem* 2013; **32:** 1288-1298.

151 Lerner M, Haneklaus M, Harada M, Grander D. MiR-200c regulates Noxa expression and sensitivity to proteasomal inhibitors. *PLoS One* 2012; **7:** e36490.

152 Howe EN, Cochrane DR, Cittelly DM, Richer JK. miR-200c targets a NF-kappaB up-regulated TrkB/NTF3 autocrine signaling loop to enhance anoikis sensitivity in triple negative breast cancer. *PLoS One* 2012; **7:** e49987.

153 Sun Y, Shen S, Liu X, Tang H, Wang Z, Yu Z *et al*. miR-429 inhibits cells growth and invasion and regulates EMT-related marker genes by targeting Onecut2 in colorectal carcinoma. *Mol Cell Biochem* 2014; **390:** 19-30.

154 Huang W, Liu H, Wang T, Zhang T, Kuang J, Luo Y *et al*. Tonicity-responsive microRNAs contribute to the maximal induction of osmoregulatory transcription factor OREBP in response to high-NaCl hypertonicity. *Nucleic Acids Res* 2011; **39:** 475-485.

155 Murray AR, Chen Q, Takahashi Y, Zhou KK, Park K, Ma JX. MicroRNA-200b downregulates oxidation resistance 1 (Oxr1) expression in the retina of type 1 diabetes model. *Invest Ophthalmol Vis Sci* 2013; **54:** 1689-1697.

156 Becker LE, Lu Z, Chen W, Xiong W, Kong M, Li Y. A systematic screen reveals MicroRNA clusters that significantly regulate four major signaling pathways. *PLoS One* 2012; **7:** e48474.

157 Zhang X, Zhang B, Gao J, Wang X, Liu Z. Regulation of the microRNA 200b (miRNA-200b) by transcriptional regulators PEA3 and ELK-1 protein affects expression of Pin1 protein to control anoikis. *J Biol Chem* 2013; **288:** 32742-32752.

158 Patel V, Hajarnis S, Williams D, Hunter R, Huynh D, Igarashi P. MicroRNAs Regulate Renal Tubule Maturation through Modulation of Pkd1. *J Am Soc Nephrol* 2012; **23:** 1941-1948.

159 Hu W, Wang X, Ding X, Li Y, Zhang X, Xie P *et al*. MicroRNA-141 represses HBV replication by targeting PPARA. *PLoS One* 2012; **7:** e34165.

160 Nagalla S, Shaw C, Kong X, Kondkar AA, Edelstein LC, Ma L *et al*. Platelet microRNA-mRNA coexpression profiles correlate with platelet reactivity. *Blood* 2011; **117:** 5189-5197.

161 Borbone E, De Rosa M, Siciliano D, Altucci L, Croce CM, Fusco A. Up-regulation of miR-146b and down-regulation of miR-200b contribute to the cytotoxic effect of histone deacetylase inhibitors on ras-transformed thyroid cells. *J Clin Endocrinol Metab* 2013; **98:** E1031-1040.

162 Li R, He JL, Chen XM, Long CL, Yang DH, Ding YB *et al*. MiR-200a is involved in proliferation and apoptosis in the human endometrial adenocarcinoma cell line HEC-1B by targeting the tumor suppressor PTEN. *Mol Biol Rep* 2014; **41:** 1977-1984.

163 Liao C, Chen W, Fan X, Jiang X, Qiu L, Chen C *et al*. MicroRNA-200c Inhibits Apoptosis in Pituitary Adenoma Cells by Targeting the PTEN/Akt Signaling Pathway. *Oncol Res* 2014; **21:** 129-136.

164 Shen LJ, He JL, Yang DH, Ding YB, Chen XM, Geng YQ *et al*. Mmu-microRNA-200a overexpression leads to implantation defect by targeting phosphatase and tensin homolog in mouse uterus. *Reprod Sci* 2013; **20:** 1518-1528.

165 You X, Liu F, Zhang T, Li Y, Ye L, Zhang X. Hepatitis B virus X protein upregulates oncogene Rab18 to result in the dysregulation of lipogenesis and proliferation of hepatoma cells. *Carcinogenesis* 2013; **34:** 1644-1652.

166 Liu Q, Tang H, Liu X, Liao Y, Li H, Zhao Z *et al*. miR-200b as a prognostic factor targets multiple members of RAB family in glioma. *Med Oncol* 2014; **31:** 859.

167 Li L, Tang J, Zhang B, Yang W, Liugao M, Wang R *et al*. Epigenetic modification of MiR-429 promotes liver tumour-initiating cell properties by targeting Rb binding protein 4. *Gut* 2014.

168 Chang L, Guo F, Wang Y, Lv Y, Huo B, Wang L *et al*. MicroRNA-200c regulates the sensitivity of chemotherapy of gastric cancer SGC7901/DDP cells by directly targeting RhoE. *Pathol Oncol Res* 2014; **20:** 93-98.

169 Peng F, Jiang J, Yu Y, Tian R, Guo X, Li X *et al*. Direct targeting of SUZ12/ROCK2 by miR-200b/c inhibits cholangiocarcinoma tumourigenesis and metastasis. *Br J Cancer* 2013; **109:** 3092-3104.

170 Szczyrba J, Nolte E, Wach S, Kremmer E, Stohr R, Hartmann A *et al*. Downregulation of Sec23A protein by miRNA-375 in prostate carcinoma. *Mol Cancer Res* 2011; **9:** 791-800.

171 Salomonis N, Schlieve CR, Pereira L, Wahlquist C, Colas A, Zambon AC *et al*. Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc Natl Acad Sci U S A* 2010; **107:** 10514-10519.

172 Xiao J, Gong AY, Eischeid AN, Chen D, Deng C, Young CY *et al*. miR-141 modulates androgen receptor transcriptional activity in human prostate cancer cells through targeting the small heterodimer partner protein. *Prostate* 2012; **72:** 1514-1522.

173 Eades G, Yao Y, Yang M, Zhang Y, Chumsri S, Zhou Q. miR-200a regulates SIRT1 expression and epithelial to mesenchymal transition (EMT)-like transformation in mammary epithelial cells. *J Biol Chem* 2011; **286:** 25992-26002.

174 Baseler WA, Thapa D, Jagannathan R, Dabkowski ER, Croston TL, Hollander JM. miR-141 as a regulator of the mitochondrial phosphate carrier (Slc25a3) in the type 1 diabetic heart. *Am J Physiol Cell Physiol* 2012; **303:** C1244-1251.

175 Braun J, Hoang-Vu C, Dralle H, Huttelmaier S. Downregulation of microRNAs directs the EMT and invasive potential of anaplastic thyroid carcinomas. *Oncogene* 2010; **29:** 4237-4244.

176 Chen Y, Xiao Y, Ge W, Zhou K, Wen J, Yan W *et al*. miR-200b inhibits TGF-beta1-induced epithelial-mesenchymal transition and promotes growth of intestinal epithelial cells. *Cell Death Dis* 2013; **4:** e541.

177 Liao X, Xue H, Wang YC, Nazor KL, Guo S, Trivedi N *et al*. Matched miRNA and mRNA signatures from an hESC-based in vitro model of pancreatic differentiation reveal novel regulatory interactions. *J Cell Sci* 2013; **126:** 3848-3861.

178 Buller B, Chopp M, Ueno Y, Zhang L, Zhang RL, Morris D *et al*. Regulation of serum response factor by miRNA-200 and miRNA-9 modulates oligodendrocyte progenitor cell differentiation. *Glia* 2012; **60:** 1906-1914.

179 Williams KC, Renthal NE, Condon JC, Gerard RD, Mendelson CR. MicroRNA-200a serves a key role in the decline of progesterone receptor function leading to term and preterm labor. *Proc Natl Acad Sci U S A* 2012; **109:** 7529-7534.

180 Cui Y, Chen J, He Z, Xiao Y. SUZ12 depletion suppresses the proliferation of gastric cancer cells. *Cell Physiol Biochem* 2013; **31:** 778-784.

181 Lin J, Liu C, Gao F, Mitchel RE, Zhao L, Yang Y *et al*. miR-200c enhances radiosensitivity of human breast cancer cells. *J Cell Biochem* 2013; **114:** 606-615.

182 Wang B, Koh P, Winbanks C, Coughlan MT, McClelland A, Watson A *et al*. miR-200a Prevents renal fibrogenesis through repression of TGF-beta2 expression. *Diabetes* 2011; **60:** 280-287.

183 Ali-Osman F, Rairkar A, Young P. Formation and repair of 1,3-bis-(2-chloroethyl)-1-nitrosourea and cisplatin induced total genomic DNA interstrand crosslinks in human glioma cells. *Cancer Biochem Biophys* 1995; **14:** 231-241.

184 Liu Y, Ding Y, Huang J, Wang S, Ni W, Guan J *et al*. MiR-141 Suppresses the Migration and Invasion of HCC Cells by Targeting Tiam1. *PLoS One* 2014; **9:** e88393.

185 Dai Y, Xia W, Song T, Su X, Li J, Li S *et al*. MicroRNA-200b is overexpressed in endometrial adenocarcinomas and enhances MMP2 activity by downregulating TIMP2 in human endometrial cancer cell line HEC-1A cells. *Nucleic Acid Ther* 2013; **23:** 29-34.

186 Xu L, Li Q, Xu D, Wang Q, An Y, Du Q *et al*. hsa-miR-141 downregulates TM4SF1 to inhibit pancreatic cancer cell invasion and migration. *Int J Oncol* 2014; **44:** 459-466.

187 Cochrane DR, Spoelstra NS, Howe EN, Nordeen SK, Richer JK. MicroRNA-200c mitigates invasiveness and restores sensitivity to microtubule-targeting chemotherapeutic agents. *Mol Cancer Ther* 2009; **8:** 1055-1066.

188 Prislei S, Martinelli E, Mariani M, Raspaglio G, Sieber S, Ferrandina G *et al*. MiR-200c and HuR in ovarian cancer. *BMC Cancer* 2013; **13:** 72.

189 Lee ST, Feng M, Wei Y, Li Z, Qiao Y, Guan P *et al*. Protein tyrosine phosphatase UBASH3B is overexpressed in triple-negative breast cancer and promotes invasion and metastasis. *Proc Natl Acad Sci U S A* 2013; **110:** 11121-11126.

190 Chang SH, Lu YC, Li X, Hsieh WY, Xiong Y, Ghosh M *et al*. Antagonistic function of the RNA-binding protein HuR and miR-200b in post-transcriptional regulation of vascular endothelial growth factor-A expression and angiogenesis. *J Biol Chem* 2012; **288:** 4908-4921.

191 McArthur K, Feng B, Wu Y, Chen S, Chakrabarti S. MicroRNA-200b regulates vascular endothelial growth factor-mediated alterations in diabetic retinopathy. *Diabetes* 2011; **60:** 1314-1323.

192 Sossey-Alaoui K, Bialkowska K, Plow EF. The miR200 family of microRNAs regulates WAVE3-dependent cancer cell invasion. *J Biol Chem* 2009; **284:** 33019-33029.

193 Tang H, Kong Y, Guo J, Tang Y, Xie X, Yang L *et al*. Diallyl disulfide suppresses proliferation and induces apoptosis in human gastric cancer through Wnt-1 signaling pathway by up-regulation of miR-200b and miR-22. *Cancer Lett* 2013; **340:** 72-81.

194 Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C *et al*. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* 2007; **131:** 1109-1123.

195 Imanaka Y, Tsuchiya S, Sato F, Shimada Y, Shimizu K, Tsujimoto G. MicroRNA-141 confers resistance to cisplatin-induced apoptosis by targeting YAP1 in human esophageal squamous cell carcinoma. *J Hum Genet* 2011; **56:** 270-276.

196 Zhu ZM, Xu YF, Su QJ, Du JD, Tan XL, Tu YL *et al*. Prognostic significance of microRNA-141 expression and its tumor suppressor function in human pancreatic ductal adenocarcinoma. *Mol Cell Biochem* 2014; **388:** 39-49.

197 Capobianco V, Nardelli C, Ferrigno M, Iaffaldano L, Pilone V, Forestieri P *et al*. miRNA and protein expression profiles of visceral adipose tissue reveal miR-141/YWHAG and miR-520e/RAB11A as two potential miRNA/protein target pairs associated with severe obesity. *J Proteome Res* 2012; **11:** 3358-3369.

198 Yu KR, Lee S, Jung JW, Hong IS, Kim HS, Seo Y *et al*. MicroRNA-141-3p plays a role in human mesenchymal stem cell aging by directly targeting ZMPSTE24. *J Cell Sci* 2013; **126:** 5422-5431.

199 Bai WD, Ye XM, Zhang MY, Zhu HY, Xi WJ, Huang X *et al*. MiR-200c suppresses TGF-beta signaling and counteracts trastuzumab resistance and metastasis by targeting ZNF217 and ZEB1 in breast cancer. *Int J Cancer* 2014; **135:** 1356-1368.

Chapter 3

Discussion and conclusions

CONTENTS

1. Discussion

Advances in miRNA research have been limited by the ability to predict biologically relevant miRNA-mRNA interactions. This is still one of the most challenging issues in this field. Overcoming this bottleneck will allow comprehensive and accurate identification of miRNA biological functions.

In this thesis an integrated framework of biochemical experimentation and bioinformatic analysis was developed to provide the first high-confidence genome-wide map of direct miR-200c target genes. The biological role of some of these novel miR-200c target genes was also investigated. Together, these results reveal how miR-200 regulates EMT and cell adhesion, which are linked to development and tumor metastasis. Importantly, this framework can be used to identify the genome-wide targets of other miRNAs. In this chapter the main advances achieved in this thesis are discussed and their inherent limitations highlighted.

1.1. Bi-miR-200c pull-down false-positive rate

The false-positive rate of the Bi-miR-200c PD was estimated to be 15.6%. This false-positive rate is similar to an estimate made for the specificity of a similarly-performed Bi-miR-34a PD, where only 1 of 11 (9%) of a random list of putative miR-34a-regulated targets was not verified[1]. The false-positive rates of bioinformatic target prediction tools such as the refined versions of TargetScan, PicTar and miRanda are higher (between 22 and 31%[2]). Studies in which miRNAs were co-immunoprecipitated with Ago proteins have much higher false-positive rates (~40-70%[3]). More-recent Ago HITS-CLIP methods for miRNA target identification have false-positive rates comparable to ours (~13–27%[3]). Given that the PD method is much simpler than Ago HITS-CLIP methods (does not require cross-linking, it is easier to perform and requires many fewer cells) it has the potential to be the method of choice for miRNA target discovery.

1.2. miR-200c-regulated pathways

The framework developed in this thesis was used to generate the first genome-wide map of miR-200c-regulated pathways. The most significantly-enriched pathways (adjusted $P<0.01$) were EGFR and MAPK signaling among thirteen significantly-enriched pathways (Figure 1d). Results show a predominance for EMT and cell adhesion-related signaling pathways that are targeted by miR-200. Some signaling pathways are involved in both EMT and cell adhesion. Specifically, EGF, sphingosine-1-phosphate (S1P), TGF-β and HGF ligands are known to induce EMT[4-6]. Additionally, MAPK signaling pathways mediate the EMT-inducing Smad-independent TGF-β response, in which TGF-β induces EMT through MAPK signaling instead of the canonical Smad signaling pathway for Smad-independent TGF-beta responses[7]. Our results also show that miR-200 regulates cell adhesion by targeting the integrin-mediated cell adhesion pathway that mediates cellular interactions with the extracellular matrix. miR-200 also regulates cell adhesion by silencing the S1P receptor signaling pathway, which regulates expression of adhesion molecules[8], and it also silences both sphingolipid and cholesterol biosynthesis pathways required by the S1P receptor signaling pathway. Three enzymes involved in amino acid metabolism, frequently altered in cancer cells, were also putative miR-200 targets.

T/B cell receptor, interleukin-3 (IL-3, mainly active in hematopoietic stem cells) and Kit signaling pathways (mainly active in hematopoietic stem cells, mast cells, melanocytes and germ cells) are not expressed in these cells, but were found to be significantly enriched by the pathway enrichment analysis algorithm due to the fact that these signaling pathways share the same downstream molecules as the EGFR and MAPK pathways.

1.3. miR-200c effect on Snail1 protein

In this thesis we found that miR-200c had no effect on Snail1 protein (Figure 15g) as confirmed by an independent study[9], even though miR-200c down-regulated *Snail1* mRNA by ~40% and a functional MRE was identified in the *Snail1* 3'UTR. This raises the question of why some targets are more responsive to a given miRNA than others. In the current miRNA action-mechanism model, miRISC silences target mRNAs mostly through mRNA degradation[10]. Therefore, genes that are transcribed into mRNAs with long half-lives are more sensitive to miRNA silencing than genes that are transcribed into mRNAs with short half-lives. Accordingly, *Snail1* mRNA is a very labile mRNA that is translated into a similarly labile protein, both with very short half-lives (*Snail1* mRNA half-life is ~15 min[11] and Snail1 protein half-life is ~25 min[12]). Therefore, the short-lived *Snail1* mRNA likely makes it less sensitive to miR-200c-mediated regulation. Therefore, no miR-200 effect on Snail1 protein was observed. Also, as previously mentioned, post-translational modifications that increase Snail1 protein half-life might counteract miR-200c-mediated down-regulation of *Snail1* mRNA. Additionally, Snail1 by itself has no transcriptional repressor activity and requires other cofactors (14-3-3 proteins, Prmt5, Ajuba, Smad3, Smad4, etc) bound together in a protein complex to exert its activity. Some of these cofactors (Ywhab/14-3-3β and Ywhag/14-3-3γ) were found in this thesis to be silenced by miR-200c. Therefore, miR-200c-mediated silencing of these cofactors might counteract Snail1 repressor activity without changes in Snail1 protein. This phenomenon of a miRNA targeting several components of a protein complex, as found in this thesis for the Zeb2 and Snail1 transcriptional repressor complexes, is observed for other miRNAs[13]. It suggests that miRNAs coordinately silence multiple proteins of the same complex as a redundant fail-proof system to alter a biological state rapidly and unequivocally.

1.4. Regulation of E-cadherin expression by Ywhag and Smad5

E-cadherin expression is a hallmark of MET[14] and this thesis shows that miR-200c-mediated silencing of Smad5 and Ywhag activates transcription from the E-cadherin promoter inserted into a reporter vector (Figure 15b). However, when *Smad5* and *Ywhag* (*14-3-3γ*) mRNAs are silenced by siRNA, endogenous E-cadherin expression does not change at the mRNA or protein levels (Figures 16d and 16e). These results suggest that the overlapping functions between members of the Smad and 14-3-3 protein families can compensate for the loss of one of the members to maintain the phenotype and homeostasis. Specifically, Zeb2 binds to several Smad proteins (Smad1, -2, -3 or -5[13]) and Snail1 binds to five of the seven 14-3-3 mammalian protein family members (14-3-3γ, ε, τ, η, β[15]). Therefore, lack of one single family member, as exemplified by our results for individual *Smad5* and *Ywhag* knock-down, might be readily compensated by another family member in order to maintain cell homeostasis. In another set of experiments, over-expression of miR-200-resistant *Smad5* slightly reduced miR-200c-mediated epithelial gene upregulation, and expression of miR-200-resistant *Ywhag* had no measurable effect (Figure 16b). This apparently minor role for *Smad5* and *Ywhag* in mediating miR-200c-epithelial gene upregulation could be due to lack of function of the over-expressed recombinant Smad5 and Ywhag proteins. Even though recombinant Smad5 and Ywhag protein over-expression was confirmed by immunoblotting analyses (Figure S2c), no functional assays were performed to confirm whether these over-expressed proteins are as active as the endogenous ones. Both over-expressed recombinant Smad5 and Ywhag proteins need to be translocated into the nucleus to regulate epithelial gene expression and this was not confirmed. Therefore, it cannot be concluded that miR-200-induced epithelial gene upregulation is indeed mediated by Ywhag. For Smad5, on the other hand, our results show some contribution to mediating miR-200-induced epithelial gene upregulation.

Considering the experimental caveats mentioned above, Smad5 might even play a stronger role in miR-200-mediated MET than our results suggest.

2. Conclusions

Based on the work within this thesis, the following conclusions can be outlined:

- The miRNA-target discovery framework developed in this thesis has the potential to be the method of choice for miRNA target discovery due to its simplicity and low false-positive rate;

- The framework developed in this thesis identified 520 novel miR-200 putative target genes and revealed that miR-200 predominantly regulates EMT and cell adhesion, processes linked to development and tumor metastasis;

- Twelve genes (*Crtap, Fhod1, Smad2, Map3k1, Tob1, Ywhag/14-3-3γ, Ywhab/14-3-3β, Smad5, Zfp36, Xbp1, Mapk12* and *Snail1*) were experimentally validated as novel miR-200c targets and their miR-200c recognition elements were precisely identified;

- We found that miR-200 induces RNA polymerase II localization and reduces Zeb2 and Snail1 transcriptional repressor complexes binding to epithelial gene promoters by silencing multiple members of these complexes;

- Three novel miR-200 target genes identified in this thesis (*Smad5*, *Ywhag* and *Crtap*) were shown for the first time to have an important role in regulating cell invasion;

Together, the results shown in this thesis make a contribution to the study of both miR-200 and EMT and will be the springboard for future mechanistic and biological investigations. Importantly, the miRNA-target discovery framework developed in this thesis could be used to identify the targets of other miRNAs and better understand their biological functions.

3. Future work

This thesis provided the first genome-wide map of miR-200c target genes and investigated the biological role of some of these novel target genes. As only some of the 520 novel miR-200c target genes identified in this thesis were experimentally validated, it would be important to experimentally validate more of these putative target genes and study their role in miR-200-mediated biology. This might unravel novel and unexpected biological functions for these genes. Also, applying the framework developed in this thesis to other members of the miR-200 family would identify the overlapping and non-overlapping roles of the several miR-200 family members.

Additional *in vitro* studies could better clarify the role of *Smad5* and *Ywhag* in mediating miR-200-induced epithelial gene upregulation. For example, co-immunoprecipitation (co-IP) experiments to identify Zeb2 and Snail1 associated-proteins could confirm whether in 4TO7 cells Smad5 and Ywhag indeed associate with Zeb2 and Snail1 transcriptional repressor complexes. Also, repeating the ChIP assays performed in this thesis to test whether silencing of *Smad5* and *Ywhag* by siRNA recapitulates miR-200-mediated reduction of Zeb2 and Snail1 binding to epithelial gene promoters would undoubtedly help to clarify the role of *Smad5* and *Ywhag* in miR-200-mediated biology.

Additional experiments would help understand the role of *Crtap* in cell invasion. Cotransfection of 4TO7 cells with an expression plasmid encoding

miR-200-resistant *Crtap* together with control miRNA or miR-200c would test whether miR-200-resistant *Crtap* can block miR-200 inhibitory effect on cell invasion, which would mean *Crtap* is required for miR-200-mediated effect on cell invasion. Also, as loss of Crtap abolishes prolyl 3-hydroxylation of type I collagen which mainly affects collagen-protein interactions[16], it would be important to test if this collagen modification is indeed altered in miR-200-transfected cells and in miR-200-expressing tumors. Also, it could be tested whether miR-200 over-expression or *Crtap* knockdown decrease the amount of decorin-bound TGF-β in the ECM, as release of TGF-β by the ECM could induce EMT in nearby cells.

Furthermore, performing *in vivo* metastasis assays using 4TO7 cells in which *Smad5, Ywhag* and *Crtap* were individually silenced by short-hairpin RNAs (shRNAs) would help understand the role of these novel target genes in mediating miR-200-induced metastasis formation.

Together these additional experiments would undoubtedly help clarify the fascinating and unexpected roles of the miR-200 family.

5. References

1 Lal A, Thomas MP, Altschuler G, Navarro F, O'Day E, Li XL *et al*. Capture of microRNA-bound mRNAs identifies the tumor suppressor miR-34a as a regulator of growth factor signaling. *PLoS Genet* 2011; **7:** e1002363.

2 Ahmadi H, Ahmadi A, Azimzadeh-Jamalkandi S, Shoorehdeli MA, Salehzadeh-Yazdi A, Bidkhori G *et al*. HomoTarget: a new algorithm for prediction of microRNA targets in Homo sapiens. *Genomics* 2013; **101:** 94-100.

3 Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009; **460:** 479-486.

4 Xue L, Su D, Li D, Gao W, Yuan R, Pang W. miR-200 Regulates Epithelial-Mesenchymal Transition in Anaplastic Thyroid Cancer via EGF/EGFR Signaling. *Cell Biochem Biophys* 2014.

5 Khan KN, Kitajima M, Hiraki K, Fujishita A, Nakashima M, Masuzaki H. Involvement of hepatocyte growth factor-induced epithelial-mesenchymal transition in human adenomyosis. *Biol Reprod* 2015; **92:** 35.

6 Milara J, Navarro R, Juan G, Peiro T, Serrano A, Ramon M *et al*. Sphingosine-1-phosphate is increased in patients with idiopathic pulmonary fibrosis and mediates epithelial to mesenchymal transition. *Thorax* 2012; **67:** 147-156.

7 Derynck R, Zhang YE. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* 2003; **425:** 577-584.

8 Kimura T, Tomura H, Mogi C, Kuwabara A, Ishiwara M, Shibasawa K *et al*. Sphingosine 1-phosphate receptors mediate stimulatory and inhibitory signalings for expression of adhesion molecules in endothelial cells. *Cell Signal* 2006; **18:** 841-850.

9 Korpal M, Ell BJ, Buffa FM, Ibrahim T, Blanco MA, Celia-Terrassa T *et al*. Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* 2011; **17:** 1101-1108.

10 Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 2010; **466:** 835-840.

11 Boettiger AN, Levine M. Rapid transcription fosters coordinate snail expression in the Drosophila embryo. *Cell Rep* 2013; **3:** 8-15.

12 Zhou BP, Deng J, Xia W, Xu J, Li YM, Gunduz M *et al*. Dual regulation of Snail by GSK-3beta-mediated phosphorylation in control of epithelial-mesenchymal transition. *Nat Cell Biol* 2004; **6:** 931-940.

13 Sass S, Dietmann S, Burk UC, Brabletz S, Lutter D, Kowarsch A *et al*. MicroRNAs coordinately regulate protein complexes. *BMC Syst Biol* 2011; **5:** 136.

14 Onder TT, Gupta PB, Mani SA, Yang J, Lander ES, Weinberg RA. Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res* 2008; **68:** 3645-3654.

15 Hou Z, Peng H, White DE, Wang P, Lieberman PM, Halazonetis T *et al*. 14-3-3 binding sites in the snail protein are essential for snail-mediated transcriptional repression and epithelial-mesenchymal differentiation. *Cancer Res* 2010; **70:** 4385-4393.

16 Morello R, Bertin TK, Chen Y, Hicks J, Tonachini L, Monticone M *et al*. CRTAP is required for prolyl 3- hydroxylation and mutations cause recessive osteogenesis imperfecta. *Cell* 2006; **127:** 291-304.